Statistical Analysis of Spatio-Temporal Veterinary Surveillance Data: Applications of Integrated Nested Laplace Approximations

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Birgit Schrödle

aus

Deutschland

Promotionskomitee

Prof. Dr. Leonhard Held (Vorsitz)
Prof. Dr. Reinhard Furrer
Prof. Dr. Andrew Hector

Zürich, 2011

Preface

On this occasion I would like to mention some people who strongly supported me during my PhD thesis.

I thank my supervisor Leonhard Held for proposing the very interesting topic and giving me the opportunity to work on it. Also, I am grateful for the deep insight into Bayesian methodology and modelling and for reading many, many provisional manuscripts very quickly.

Furthermore, the financial support by the Swiss federal veterinary office (BVET) is gratefully acknowledged. In particular, I thank Jürg Danuser and Heinzpeter Schwermer for providing the Swiss registry data, raising interesting questions and sharing their knowledge on veterinary legislation with me.

Many thanks go to all my colleagues at the Biostatistics division for an extremely inspiring working environment, all the linux and latex support, the help with proofreading and the encouragement at conferences.

All this work would not have been possible without the great INLA support by Håvard Rue. I am very thankful that I could work in Trondheim at two occasions and that I experienced the great hospitality of Håvard and the NTNU statistics group there. Thank you for all the secret INLA options.

I thank the referees and the dissertation committee for reviewing this thesis.

Coming to Zurich was one of the best decisions in my life. Besides work, additional encouragement was given to me by all my friends and, especially, my flatmate Veronique. As there are too many great experiences to be mentioned, here are just a few impressions: Acapulco, Aletsch-Gletscher, Allgäu Statistician weekend, Barbecue Paradise Fritschiwiese, Europapark, Fokker 100, Fanmeile Fussball-WM, Himbeer Daiquiri, Jungfraujoch, Lindy Hop, München-Freiburg-Tilburg-Zürich, playing at Bitli's, snowshoeing, statistical hiking tours, super biker girl, ZYP49, ...

Finally, I thank my family: Mama and Simon, and Reto, for all their love, support, belief and much more.

Zurich, April 2011

Birgit Schrödle

Zusammenfassung

Die Überwachung von Tierkrankheiten ist eine wichtige Aufgabe nationaler Veterinärämter. Zentrale Ziele sind die Vermeidung der Ausbreitung und, im Fall von Zoonosen, Übertragung von Krankheiten vom Tier auf den Menschen. Weit verbreitet ist eine sogenannte passive Überwachung, bei der diagnostizierte Fälle gemeldet werden müssen. Allerdings sind solche Daten häufig durch verspätete bzw. unterbleibende Meldungen verzerrt.

Das Schweizer Bundesamt für Veterinärwesen (BVET) führt seit 1991 ein Register mit circa 80 meldepflichtigen Krankheiten. Dieses enthält neben der Wochennumber der Diagnose auch eine Angabe darüber, in welcher der 184 administrativen Regionen ein Fall auftrat. Zusätzliche Daten kommen aus dem Fürstentum Liechtenstein. Ziel dieser Dissertation ist es, basierend auf den Charakteristika einer Krankheit Ansätze zur statistischen Modellierung räumlich-zeitlicher Muster vorzuschlagen und diese auf ausgewählte Datensätze aus der BVET Datenbank anzuwenden. Werden durch eine solche Modellierung z.B. Regionen mit hoher Inzidenz einer Krankheit identifiziert, so können gezielte Massnahmen folgen. Ein weiterer Schwerpunkt liegt auf der Vorstellung benutzerfreundlicher Software und verfügbarer Modellwahlkriterien.

Räumlich-zeitliche Zähldaten werden häufig mit hierarchischen Bayesianischen Modellen analysiert. Als Inferenzmethode stellen wir integrierte, genestete Laplace Approximationen (INLA) vor und zeigen ihre vielseitige Anwendbarkeit für Raum-Zeit-Modelle auf. Hohe Benutzerfreundlichkeit ist durch die frei verfügbare INLA Software gewährleistet. Neben dem Devianz-Informations-Kriterium diskutieren wir vorhersagebasierte Scores, die von INLA zur Modellwahl und -validierung bereitgestellt werden. Solche Scores erweisen sich sowohl zur Bewertung von kreuzvalidierten als auch von Ein-Schritt-Vorhersagen als nützlich.

Zunächst analysieren wir aggregierte Regionsdaten von Krankheiten mit konstantem, endemischen Auftreten. Neben der Modellierung von räumlicher Autokorrelation diskutieren wir in einer Studie zu Coxiellose beim Rind das Einbeziehen eines linearen Zeittrends. Desweiteren behandeln wir die geeignete Berücksichtigung (linear, nichtparametrisch) einer regionsspezifischen Kovariable (ökologische Regression). In einer Analyse von Salmonellose Fällen beim Rind schlagen wir einen nichtparametrischen Zeittrend vor und diskutieren Optionen bezüglich dessen Modellierung. Als weiteren Schwerpunkt beschreiben wir die vielseitige inhaltliche Bedeutung räumlich-zeitlicher Interaktionen und zeigen Kriterien bezüglich ihrer Identifizierbarkeit auf. Um das Meldeverhalten bei Boviner Virus Diarrhoe in Bezug auf die Zugehörigkeit einer Region zu einem Schweizer Kanton zu analysieren, erweitern wir diese Modelle um eine gröbere, kantonale Ebene. Ein Vergleich mit ausschliesslich regionalen Modellen durch kreuzvalidierte Scores weist ein unterschiedliches Meldeverhalten in einzelnen Schweizer Kantonen nach.

Für Blauzungenkrankheit (BT) wurde 2008/09 ein aktives Überwachungs- und Impfprogramm in der Schweiz durchgeführt. Wir führen eine Regression durch, die neben einem zweidimensionalen, räumlichen Effekt auf einem regulären Gitter auch individuelle Daten zu Impfung, Überwachungsform und Höhenlage jeder Tierhaltung in Beziehung zum Auftreten von BT setzt. Die Ergebnisse deuten darauf hin, dass eine Impfung das Risiko einer BT Infektion verringert.

Zur Modellierung von Krankheiten mit lokalen Ausbrüchen schlagen wir ein vektorautoregressives Modell für multivariate Zeitreihen vor. Wir zeigen auf, wie Daten über Netzwerke zwischen einzelnen Regionen direkt ins Modell einzubezogen werden können. Für Coxiellose Fälle beim Rind weisen wir nach, dass eine räumlich-zeitliche Ausbreitung zwischen benachbarten Regionen und durch den Tierhandel stattfindet. Für diese Fallstudie zeigen wir mithilfe vorhersagebasierter Scores auf, dass eine solche parametergetriebene Modellierung eine bessere prädiktive Güte bietet als sogennante beobachtungsgetriebene Modelle, in denen tatsächlich beobachtete Fälle zum vorherigen Zeitpunkt einen Ausbruch steuern.

Abstract

The surveillance of animal diseases is an important task of national veterinary authorities. Major aims are the prevention of disease spread and, for zoonoses, the transmission of diseases from animals to humans. Monitoring is mostly done by passive surveillance, where laboratory confirmed cases have to be reported. However, such data are often biased due to reporting delay or underreporting.

Since 1991 the Swiss federal veterinary office (BVET) collects data on about 80 notifiable diseases. The week number of diagnosis and the location within one of 184 administrative regions is known for each case. Additionally, data from the Principality of Liechtenstein are available. The aim of this dissertation is to propose approaches for the statistical modelling of spatio-temporal patterns based on the specific characteristics of a disease and to apply them to selected data from the BVET database. If, *e.g.*, regions with a high disease incidence are identified by such a model, appropriate control measures can be initiated. A further emphasis is on the presentation of user-friendly software and available model choice criteria.

Spatio-temporal count data are often analyzed using hierarchical Bayesian models. For inference we propose integrated nested Laplace approximations (INLA) and show their versatile applicability as regards space-time modelling. High usability is guaranteed by freely available INLA software. Along with the deviance information criterion we discuss predictive scores, which are provided by INLA for model choice and criticism. Such scores turn out to be useful for the evaluation of cross-validatory as well as one-step-ahead forecasts.

We begin with an analysis of aggregated regional data of diseases with constant, endemic risk. In addition to modelling spatial autocorrelation we describe the inclusion of a linear time trend for a case study on Coxiellosis in cows. Furthermore, we discuss the appropriate specification (linear, nonparametric) of a region-specific covariate. For an analysis of Salmonellosis cases in cows we propose a nonparametric time trend and discuss various modelling options. A further emphasis is on the versatile interpretation of spatio-temporal interaction terms and the derivation of criteria to guarantee their identifiability. To analyze case reporting of Bovine Viral Diarrhoea concerning the affiliation of a region to a Swiss canton, we expand these models by a coarser, cantonal grid. A comparison with exclusively regional models using cross-validated scores shows a biased case reporting in several Swiss cantons.

An active surveillance and vaccination program was launched for Bluetongue (BT) in 2008/09 within Switzerland. We perform a regression which assesses the association between individual information on vaccination, surveillance and altitude and the occurrence of BT for each farm. Additionally, a two-dimensional location effect on a regular lattice is included in the model. The results indicate that a vaccination reduces the risk of a BT infection.

We propose a vector-autoregressive model for multivariate time series to model diseases with local outbreaks. Furthermore, we show how information on networks between regions can directly be related to observed disease counts. Using this methodology, a spatio-temporal spread of Coxiellosis in cows between neighbouring regions and by cattle trade is detected. Comparing one-step-ahead predictive scores it turns out that, for this case study, such a parameter-driven approach exhibits a better predictive performance than so-called observation-driven models, where actually observed previous cases govern the infection mechanism.

Thesis outline

Introduction

Paper I: **Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA**

Leonhard Held, Birgit Schrödle & Håvard Rue

Pre-print version of a paper published in *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir, Thomas Kneib & Gerhard Tutz (eds.)*, 2010, Physica-Verlag, Heidelberg, ISBN 978-3-7908-2412-4.

Paper II: **A primer on disease mapping and ecological regression using INLA**

Birgit Schrödle & Leonhard Held

Pre-print version of a paper published in *Computational Statistics*, 2011, **26** (6), 241-258. The original publication is available at `http://www.springerlink.com/content/n158p80372434r7h/fulltext.pdf`.

Paper III: **Spatial analysis of bluetongue cases and vaccination of Swiss cattle in 2008 and 2009**

Katriina Willgert, Birgit Schrödle & Heinzpeter Schwermer

Paper published in *Geospatial Health*, 2011, **5** (2), 227-237.

Paper IV: **Spatio-temporal disease mapping using INLA**

Birgit Schrödle & Leonhard Held

Paper published in *Environmetrics*, 2010, DOI: 10.1002/env.1065.

Paper V: **Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: A case-study**

Birgit Schrödle, Leonhard Held, Andrea Riebler & Jürg Danuser

Paper published in *Journal of the Royal Statistical Society - Series C*, 2011, **60** (2), 261-279.

Paper VI: **Assessing the impact of network data on the spatio-temporal spread of infectious diseases**

Birgit Schrödle, Leonhard Held & Håvard Rue

Paper in revision for *Biometrics*.

Appendix A: **Analyzing veterinary surveillance data: Approaches to model the relationship between disease incidence and cattle trade**

Birgit Schrödle, Leonhard Held & Michaela Paul

Extended abstract published in the Proceedings of the 25^{th} *International Workshop on Statistical Modelling, Glasgow, UK, 2010.*

Introduction

The prevention of animal diseases is important for many different reasons. First of all, the animal itself has to be protected and the spread of a disease must be restrained to guarantee a healthy animal population. Furthermore, the prevalence of a disease can cause extensive economic losses. An example is Bovine Viral Diarrhoea (BVD) in Swiss cows, which until recently caused a damage of several million Swiss francs every year (BVET, 2006). The disease was eradicated by an immense program in Switzerland in 2008 (Presi and Heim, 2010; Zimmerli *et al.*, 2010). Several diseases, so-called zoonoses, can also threaten human life and are often transmitted by food. A prominent example is Bovine Spongiform Encephalopathy (BSE). Molecular, temporal and spatial data strongly suggest an association between BSE and the new variant Creutzfeldt-Jakob disease (Cohen-Sabas *et al.*, 2004). The *Salmonella* bacterium is one of the most common causes of foodborne diarrhoeal diseases worldwide (Wegener *et al.*, 2003). Furthermore, harmful bacteria which are excreted by cows, goats and sheep in their milk and feces can be transmitted to humans by airborne infection. For example, the *Coxiella burnetii* bacterium can lead to Q-Fever in humans (Raoult *et al.*, 2005).

Hence, the surveillance of animal diseases by state institutions is of great importance and mostly enshrined in law. Surveillance means to collect, manage, analyze, interpret and report information about the status of diseases in a population (Buehler, 2008). It is an ongoing, dynamic process and supports decision-making (Stärk, 1996). However, the national authorities have to find a reasonable trade-off between information needs and feasibility in the data collection (Buehler, 2008). As a consequence, we distinguish active and passive surveillance. Active surveillance means that the authorities initiate programs to collect data. Passive surveillance means that veterinarians or laboratories must report cases of notifiable diseases to the veterinary authorities. The determined level of responsibility for reporting often depends on the characteristics of a disease. Most diseases are monitored by passive surveillance due to its relative simple implementation and limited investment costs. The normal passive surveillance is often replaced by an active phase once an outbreak has been detected (Giesecke, 2002).

The analysis of spatio-temporal registry data offers several benefits. First of all, it is important to monitor disease incidence over time, as the final goal of every surveillance program should be to control and ultimately to eradicate the disease from the population (Keeling and Rohani, 2008). Furthermore, collective reporting at a national level makes it possible to see the full picture of incidence instead of only single cases observed by a veterinarian (Giesecke, 2002). Stroup *et al.* (2004) state that surveillance data can help to learn about the natural history of a disease. When outbreaks or spatio-temporal disease clusters are suspected, information about the success of interventions can be provided and can influence the direction of veterinary legislation. Statistical models can be a powerful tool to optimize the use of limited resources or to target control measures more efficiently. As a further issue, it is important to prevent the spread of infections. Analyses of, *e.g.*, trading networks tend to be very powerful tools for understanding the transmission of a disease (Keeling and Rohani, 2008). A possible consequence are short-term trade restrictions.

However, the analysis of routinely collected surveillance data poses some problems. It is known that results obtained from case reporting data can potentially be heavily biased due to limited case detection or low reporting motivation (Doherr and Audige, 2001; Salman *et al.*, 2003; Bernet, 2005). For some diseases farmers might even be afraid of the consequences of a detected case (Cohen-Sabas *et al.*, 2004). Hence, underreporting is potentially present in the data. Furthermore, there might be a time delay, as the date in the database usually corresponds to the day the confirmation form was received and not to the day of diagnosis or

onset of the disease. Some subclinical diseases are only detected by screening tests which are more frequently conducted on large farms. Regional health authorities possibly have different prevention priorities. When a longer timespan is observed, changes in the laboratory practice or interventions might have taken place. Further sources of bias are described in Giesecke (2002).

In Switzerland, about 80 diseases are denoted as “notifiable" by the Swiss Federal Veterinary Office (BVET). People who own, trade or care for animals are bound by law to report the incidence of such a disease or of symptoms to a veterinarian. All notifiable diseases are listed in the federal regulations for animal diseases and the aim of prevention is stated for each disease separately. In general, the notifiable diseases are split into four groups: “Highly contagious diseases" are diseases which quickly spread over country borders like the Foot-and-Mouth disease. A second group are “diseases to be eradicated" in Switzerland, *e.g.* BSE and BVD, which are currently not observed in Switzerland. This is monitored by active surveillance. The aim for “diseases to be fought", *e.g.* Salmonellosis, is to reduce the economic and health consequences. The least responsibility for reporting is for “diseases to be supervised" such as Coxiellosis. Here, the goal is just to collect the number of cases over time and monitor their incidence. The usual reporting chain in Switzerland is shown in Figure 1. Note that Switzerland

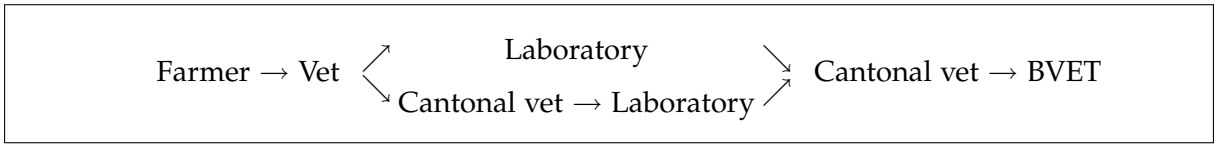


Figure 1: Reporting chain in Switzerland

is a confederation of 26 cantons which have their own veterinary authorities (“cantonal vet"). Each canton consists of one or more administrative regions. A suspicious case is reported by a farmer to a veterinarian. In the follow up each case must be confirmed by a laboratory and reported to the cantonal veterinarian. This person finally hands on a confirmed case to the BVET. As mentioned above, a reporting bias can be introduced at all stages of this reporting chain. Confirmed cases are digitally collected and archived. The basic unit of a case is a herd, which is marked as infected if at least one diseased animal was detected.

For the analyses in this dissertation the reported and approved cases of animal diseases in Switzerland are available since 1991 until present and can be assigned to one of 184 Swiss regions or the Principality of Liechtenstein. Such data are usually called aggregated data. In a few cases active surveillance measures are applied when an outbreak of the disease is detected. A prominent example is the occurrence of Bluetongue serotype 8 (BT) in Switzerland in 2008 and 2009. A large, national vaccination programme was launched subsequently.

A further important issue is that the existing cattle trade network poses a potential risk of disease transmission as it serves as a contact node between infected herds. The ease of transportation can result in the spread of a disease over long distances (Fèvre *et al.*, 2006). A well-known example is the association between cattle trade from infected areas and the breakdown of herds during the Bovine Tuberculosis epidemic in Great Britain (Gilbert *et al.*, 2005). As it is mandatory in Switzerland to notify all cattle trade via an on-line system, the cattle trade between Swiss regions can be used as a source of information for the spatio-temporal spread of diseases.

The statistical focus of this thesis is on the spatio-temporal modelling of bovine diseases, namely BVD, Salmonellosis, Coxiellosis and BT in Swiss cows, and the detection of spatial and temporal clusters. Furthermore, methods for assessing the impact of cattle trade on the

spatio-temporal spread of Coxiellosis will be proposed. Depending on the aim of the analysis, the etiology of the disease and the spatial resolution of the available data, different modelling approaches are taken. A feature that all proposed models have in common is that they fit into the class of so-called latent Gaussian models. Within the last two decades, Bayesian inference in latent Gaussian models has mainly been implemented using Markov chain Monte Carlo (MCMC) techniques (Knorr-Held, 2000; Frühwirth-Schnatter and Wagner, 2006). However, here we use integrated nested Laplace approximations (INLA) which are a recent approach for approximate Bayesian inference in latent Gaussian models (Rue *et al.*, 2009). The fascinating fact is that all later mentioned, complex models within this dissertation can be implemented within the INLA framework using freely available `R` software (R Development Core Team, 2005). A further emphasis of this thesis is on appropriate model choice criteria and their computation using INLA.

This introduction is structured as follows. In Section 1 latent Gaussian models and their properties are reviewed. Furthermore, the INLA framework is briefly introduced and its limitations are discussed. Spatial and spatio-temporal models for surveillance data and the analysis of networks are discussed in Section 2. Every subsection contains references to the respective papers and some example `R-INLA` code for illustration. Model choice in hierarchical Bayesian models using INLA is discussed in Section 3.

1 Integrated nested Laplace approximations

This chapter deals with INLA as a tool for Bayesian inference in latent Gaussian models.

1.1 Gaussian Markov random fields

The following section will introduce Bayesian hierarchical models and Gaussian Markov random fields as an important building block of the INLA approach.

Bayesian hierarchical models So-called Bayesian hierarchical models are attractive as they provide a unified approach to data analysis (Banerjee *et al.*, 2004). They are usually characterized by three stages of observations and parameters. The first stage consists of distributional assumptions for the observations. For example, if we observe disease counts y_i ($i = 1, \ldots, I$) for I geographic regions within a pre-specified time period, we may assume that y_i is Poisson distributed with rate λ_i. The parameter λ_i denotes the relative risk for a disease case in region i. We assume that the y_i's are conditionally independent given all λ_i's. On the second stage we define a prior model for the λ_i's or, more often, a specific transformation of them. In the Poisson case it is common to use $\log(\lambda_i) = \eta_i$. The variable η_i is called the linear predictor and is usually an additive term of unknown random components (Fahrmeir and Tutz, 2001). These can be of different types, *e.g.* spatial random effects and linear or smooth effects of covariates. High flexibility can be obtained by assigning Gaussian priors to all components of the linear predictor. Such models are also called latent Gaussian models (Rue *et al.*, 2009). The third stage consists of prior distributions for unknown hyperparameters $\theta_1, \ldots, \theta_R$, which typically are variances or correlations for random effects within $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_I)^T$. The priors for the hyperparameters are not necessarily Gaussian. Latent Gaussian models are widely used in the statistics literature, *e.g.* in structural time-series analysis, analysis of longitudinal and survival data and semiparametric regression (Fahrmeir and Tutz, 2001; Gelman *et al.*, 2004). The setting is especially appealing for modelling spatial and spatio-temporal data (Besag *et*

al., 1991; Banerjee *et al.*, 2004).

Gaussian Markov random fields and conditional independence We can form a large vector $\boldsymbol{x}$, which consists of the linear predictor vector $\boldsymbol{\eta}^T$ and all its additive components. As the η_i's are on the first I positions in the vector $\boldsymbol{x}$, each observation y_i depends directly only on the corresponding ith element x_i in $\boldsymbol{x}$. Furthermore, since Gaussian priors are assigned to all components of $\boldsymbol{x}$ as mentioned above, the vector $\boldsymbol{x}$ is also Gaussian and forms a so-called Gaussian Markov random field (GMRF) (Rue and Held, 2005). GMRFs are sometimes called conditional autoregressions since they fulfill conditional independence or so-called Markov properties (Besag and Kooperberg, 1995). Let $\pi(\cdot|\cdot)$ denote a conditional density of its arguments in the following. Two variables x_i and x_j ($i \neq j$) are called conditionally independent given $\boldsymbol{x}_{-ij}$, if $\pi(x_i, x_j|\boldsymbol{x}_{-ij}) = \pi(x_i|\boldsymbol{x}_{-ij}) \cdot \pi(x_j|\boldsymbol{x}_{-ij})$. This can be written as $x_i \perp x_j|\boldsymbol{x}_{-ij}$. The vector $\boldsymbol{x}_{-ij}$ contains all components of $\boldsymbol{x}$ except for the ith and jth. The conditional independence between two components x_i and x_j of a GMRF can directly be read off from its so-called precision matrix $\boldsymbol{Q}$. It holds that

$$x_i \perp x_j|\boldsymbol{x}_{-ij} \;\Leftrightarrow\; Q_{ij} = 0.$$

More formally, we can define a random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ as GMRF with mean $\boldsymbol{\mu}$ and a positive definite precision matrix $\boldsymbol{Q}$, if its density has the form

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

The covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{Q}^{-1}$ of the GMRF is the inverse of the precision matrix. At the second stage within a hierarchical model, GMRFs provide a flexible tool to model the dependence between latent effects and thus, implicitly, the dependence between the observed data (Rue and Martino, 2007).

Sparse matrix algorithms The sparseness of the precision matrix $\boldsymbol{Q}$ due to the Markov properties is the key to the fast computation of integrated nested Laplace approximations discussed below. Rue (2001) and Rue and Held (2005) discuss elaborate numerical algorithms for sparse matrices from a statistical perspective and how to apply them to GMRFs. In general, numerical methods for sparse matrices are much quicker than for dense matrices (Rue and Held, 2005, Section 2.3).

Many of these algorithms are based on the Cholesky decomposition of $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^T$. Here, $\boldsymbol{L}$ is the lower triangular Cholesky triangle. A sparse $\boldsymbol{Q}$ allows for fast factorization and inherits this property to $\boldsymbol{L}$. Further simplification of the computations can be achieved by re-ordering $\boldsymbol{Q}$ to decrease the number of non-zero terms in $\boldsymbol{L}$. To facilitate this, imagine the structure of the GMRF as an undirected graph with n vertices, where there is an edge between the vertices i and j, if the (i,j)-th entry in the precision matrix $\boldsymbol{Q}$ is different from 0. The sparseness of $\boldsymbol{L}$ depends heavily on the ordering of the indices of the GRMF $\boldsymbol{x}$. Optimal permutations of the graph can be achieved by nested dissection or converting the precision matrix into a band matrix (Rue and Held, 2005, Section 2.4.2).

The variances of the GMRF which are the diagonal elements of $\boldsymbol{\Sigma} = \boldsymbol{Q}^{-1}$ are only implicitly known. However, they are needed for the computations discussed in Section 1.2. The precision matrix $\boldsymbol{Q}$ can be formally inverted, but its dimension is typically too large. A very efficient algorithm for computation of the marginal variances Σ_{ii} and covariances Σ_{ij} based on the

Cholesky factor $\boldsymbol{L}$ is described in Rue and Martino (2007).
The open-source library `GMRFLib` includes an efficient implementation of all needed algorithms for GMRFs (Rue and Held, 2005, Appendix B).

Intrinsic GMRFs and linear constraints So-called intrinsic GMRFs (IGMRFs) play an important role in hierarchical models. The specific property of IGMRFs is that they are improper, *i.e.* their precision matrix $\boldsymbol{Q}$ is not of full rank (Rue and Held, 2005, Section 3). The central result in Rue and Held (2005, Section 3.2) is that the density $\pi^*(\boldsymbol{x})$ of an IGMRF with dimension n and precision matrix of rank $n-k$ can be written as the density of a proper GMRF under k linear constraints: $\pi(\boldsymbol{x}|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{a})$. Here, $\boldsymbol{A}$ denotes a $k \times n$ matrix of rank k and $\boldsymbol{a}$ a $k \times 1$ vector. Hence, the usual strategy to obtain an estimate of an IGMRF is to compute an unconstrained $\boldsymbol{x}$ first and then to adjust it for linear constraints using

$$\boldsymbol{x}^* = \boldsymbol{x} - \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{a}). \tag{1}$$

This is called conditioning by kriging (Rue and Held, 2005, formula (2.30)). The algorithm to obtain marginal variances is still valid for a constrained GMRF, but will not allow a fast computation, as the linear constraints destroy the sparseness of $\boldsymbol{Q}$. However, similar to the transformation of $\boldsymbol{x}$ into $\boldsymbol{x}^*$, the unconstrained covariance matrix $\boldsymbol{\Sigma}$ can be translated into the constrained covariance matrix using

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma} - \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1},$$

see Rue and Martino (2007).

Gaussian approximation An important building block for the use of INLA is the so-called Gaussian approximation. According to Knorr-Held and Rue (2002), the full conditional distribution of a zero mean GMRF $\boldsymbol{x}$ in a latent Gaussian model with hyperparameter vector $\boldsymbol{\theta}$ can be written as

$$\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_{i=1}^{I}\log\pi(y_i|x_i)\right).$$

As noted above, y_i is conditionally independent of $\boldsymbol{x}_{-i}$ given x_i. This density is non-normal, but can often be well approximated by a Gaussian distribution $\widetilde{\pi}_G$ (Rue and Held, 2005, Section 4.4.1). The approximation $\widetilde{\pi}_G$ is found by matching the mode and the curvature at the mode of the full conditional density (Rue *et al.*, 2009), resulting in

$$\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q}+\mathrm{diag}(\boldsymbol{r}))(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

Here, $\boldsymbol{\mu}$ is the mode of $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ and is computed iteratively using a Newton-Raphson algorithm (Fahrmeir and Tutz, 2001). If linear constraints are involved, $\boldsymbol{\mu}$ is corrected in each iteration using (1). The term $\boldsymbol{Q}+\mathrm{diag}(\boldsymbol{r})$ denotes the precision matrix of the approximate density. The components of the vector $\boldsymbol{r}$ are derived from the second order terms in the Taylor expansion of $\sum_{i=1}^{I}\log\pi(y_i|x_i)$ at the modal value $\boldsymbol{\mu}$. Note that $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ also depend on $\boldsymbol{\theta}$.

1.2 Integrated nested Laplace approximations

We briefly discuss the INLA approach in the following section. Since it is impossible to describe all details involved here, we refer to Rue *et al.* (2009, incl. discussion) for a compre-

hensive description.

General idea Often the main goal of an analysis of latent Gaussian models is to compute posterior marginals

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{x}_{-i}} \pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{x}_{-i} d\boldsymbol{\theta} \tag{2}$$

for each component x_i and sometimes also posterior marginals for the hyperparameters θ_r. The INLA approach uses the fact that the desired posterior marginal (2) can be re-written as

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y}) \; \pi(\boldsymbol{\theta}|\boldsymbol{y}) \; d\boldsymbol{\theta}.$$

This integral is approximated by the finite sum

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\boldsymbol{\theta}_k, \boldsymbol{y}) \; \widetilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y}) \; \Delta_k, \tag{3}$$

where $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ and $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ denote approximations of $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, respectively. Finally, the sum is evaluated at support points $\boldsymbol{\theta}_k$ by numerical integration using appropriate weights Δ_k.

Approximation of $\boldsymbol{\pi(\theta|y)}$ Factorizing $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \times \pi(\boldsymbol{\theta}|\boldsymbol{y}) \times \pi(\boldsymbol{y})$ it follows that $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ can be approximated by

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{4}$$

where the denominator $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ denotes the Gaussian approximation of $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ with mode $\boldsymbol{x}^*(\boldsymbol{\theta})$ for a given $\boldsymbol{\theta}$. This ratio corresponds to the Laplace approximation introduced by Tierney and Kadane (1986) and is mainly needed to integrate out the uncertainty with respect to $\boldsymbol{\theta}$ via (3). For this purpose it is not necessary to represent the approximation parametrically, but only to locate its mode and explore the shape of the probability mass to find appropriate evaluation points $\boldsymbol{\theta}_k = (\theta_{k,1}, \ldots, \theta_{k,R})^T$ for (3). The mode $\boldsymbol{\theta}^*$ of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is obtained using a quasi-Newton method to optimize $\log \widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ (Dennis and Schnabel, 1996). The shape of the probability mass can either be assessed by a step-wise exploration algorithm or using classical design strategies.

Approximation of $\boldsymbol{\pi(x_i|\theta, y)}$ For the approximation of $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ three strategies can be employed: A Gaussian, a full Laplace and a simplified Laplace approximation. The Gaussian approximation as introduced in Section 1.1 is computationally most convenient. Each marginal $\widetilde{\pi}_G(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ can simply be derived from $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, *i.e.* $\widetilde{\pi}_G(x_i|\boldsymbol{\theta}, \boldsymbol{y}) = N(x_i; \mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta}))$, where $\mu_i(\boldsymbol{\theta})$ is the mean of the Gaussian approximation and $\sigma_i^2(\boldsymbol{\theta})$ is the corresponding marginal variance. However, there can be errors in the location of the posterior marginals, errors due to the lack of skewness, or both (Rue and Martino, 2007). A time-consuming, but typically very exact alternative is to use another Laplace approximation for $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, similar to (4):

$$\widetilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}_{-i}|x_i, \boldsymbol{\theta}, \boldsymbol{y})}|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i, \boldsymbol{\theta})}. \tag{5}$$

The evaluation of this approximation is computationally very demanding, since the denominator must be recomputed for each x_i and $\boldsymbol{\theta}$. Hence, several numerical tricks and simplifications must be applied to make this approximation computationally feasible. A simplified representation of the final result is

$$\widetilde{\pi}_{LA}(x_i|\boldsymbol{\theta},\boldsymbol{y}) \propto N(x_i;\mu_i(\boldsymbol{\theta}),\sigma_i^2(\boldsymbol{\theta})) \times \exp(\text{cubic spline}(x_i)).$$

The so-called simplified Laplace approximation $\widetilde{\pi}_{SLA}(x_i|\boldsymbol{\theta},\boldsymbol{y})$ is less expensive than the full Laplace approximation with only a slight loss of accuracy. Here, a series expansion of the Laplace approximation (5) around $x_i = \mu_i(\boldsymbol{\theta})$ is performed. This approach allows to correct the Gaussian approximation for location and skewness.

Further issues It has been shown for a variety of applications that there is excellent agreement between MCMC and INLA results for the simplified and full Laplace approximation. As examples see Riebler *et al.* (2010) for an application of INLA to multivariate age-period-cohort models and Paper V for nonparametric spatio-temporal models.

The strength of INLA is that all involved analyses are implemented in ready-to-use software which can be run within a powerful `R` environment (R Development Core Team, 2005). An `R` package `INLA` (called `R-INLA` in the following) can be freely downloaded from `www.r-inla.org`. Hence, inference for complex hierarchical models is accessible for the end-user.

1.3 Limitations of INLA

However, some limitations of the INLA approach must be considered.

Hyperparameters Two restrictions concerning the hyperparameters $\boldsymbol{\theta}$ exist. The first requirement is that the number of hyperparameters $|\boldsymbol{\theta}|$ is small (< 6) or at least moderate (6-12). This is due to the fact that integrating out the hyperparameters is computationally expensive, *e.g.* exponential with respect to $|\boldsymbol{\theta}|$ for the step-wise exploration algorithm mentioned above. Numerical integration as in (3) becomes infeasible in higher dimensions. Less costly approaches based on classical design strategies exist, but the results are less accurate (Rue *et al.*, 2009, Section 6.5). The computation time for different strategies is compared for a spatio-temporal example in Paper II. However, there are models where a large number of hyperparameters is required, *e.g.* additive genetic models which are used in evolutionary biology (Steinsland and Jensen, 2010). Here, multiple traits of an animal are analyzed simultaneously, each of them having a genetic and environmental random effect. When correlations of genetic and environmental effects between traits are considered as well, the number of hyperparameters grows quite quickly.

Furthermore, it might be desirable to put additional structure on the hyperparameters. For example, Nobre *et al.* (2005) propose a spatio-temporal model for mapping the Malaria incidence in Pará, Brazil. They allow the variance parameters itself to vary over time by imposing a random walk prior on them. Concerning spatial variation of the variance parameters Schmid (2011) proposes a locally adaptive spatio-temporal smoothing approach for modelling perfusion cardiovascular Magnetic Resonance Imaging (MRI). Here, the adaptive smoothing allows to retain sharp features and borders of myocardial tissue areas.

Computation time In general, INLA provides very precise approximations to posterior quantities within seconds. However, some issues concerning spatio-temporal models can slow

down the approximation routine a lot. Firstly, as indicated above integrating out the uncertainty with respect to the hyperparameters can take long in specific settings. An example are generalized dynamic models as proposed to assess the impact of network data on the spatio-temporal spread of infectious diseases in Section 2.3 and Paper VI. For a first exploratory analysis of such models it is advisable to use an empirical Bayes approach, which uses only the mode $\boldsymbol{\theta}^*$ of the marginal hyperparameter posterior (Rue *et al.*, 2009, Section 6.5).

Secondly, the nonparametric spatio-temporal models discussed in Section 2.1 and applied in Papers IV and V require the inclusion of many linear constraints on the latent Gaussian field, as shown in Section 1.1. Such constraints slow down the inference process a lot if the simplified or Laplace approximation are used.

Concavity The use of INLA is limited to models where the log-likelihood is concave with respect to the linear predictor. Being concave means that the Fisher information matrix of unknown quantities in the linear predictor is positive definite. If this property is not fulfilled, Newton-Raphson algorithms involved in several steps of the INLA approach, *e.g.* the Gaussian approximations, do not converge to a local maximum (Dennis and Schnabel, 1996).

As an illustrative example we consider a log-linear Poisson model, *i.e.* a generalized linear model for count data where the mean λ is modelled via the log-link

$$y \sim \text{Po}(\lambda = \exp(\eta))$$

(Fahrmeir and Tutz, 2001, Section 2.2.1). For simplification, we assume that η is an intercept. Wedderburn (1976) shows that the log-likelihood of this model is always concave with respect to η. However, if we propose a model

$$y \sim \text{Po}(\lambda = c + \exp(\eta)), \tag{6}$$

where an additive constant c is introduced, a different situation arises. Note that this model is a simplified version of model (11) discussed later. As only one unknown parameter is involved (η), concavity is guaranteed, if the Fisher information (the negative second derivative of the log-likelihood) is positive for all values of η. However, Figure 2 shows that the Fisher information (right) can be negative for some configurations and that the log-likelihood (left) is not concave for all values. Modifications to obtain a so-called globally convergent Newton

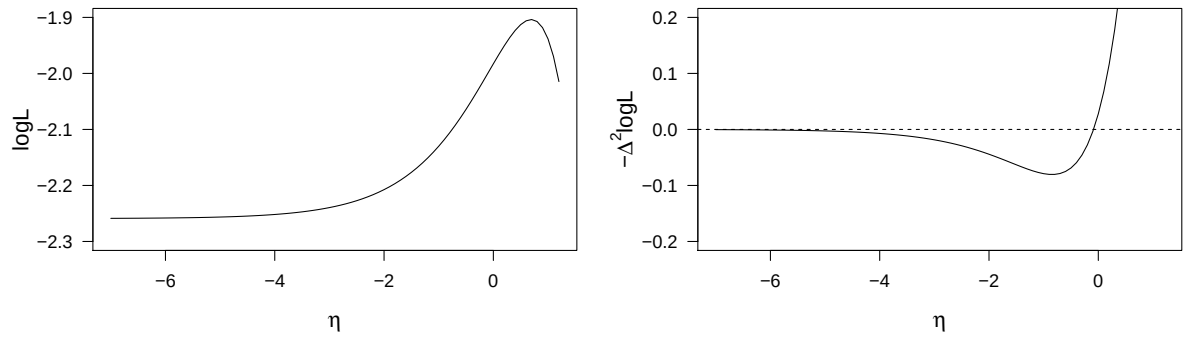


Figure 2: The log-likelihood (left) and its negative second derivative (right) for model (6) for $y = 7$, $c = 5$ and varying η

algorithm by, *e.g.*, forcing the Fisher information matrix to be positive definite exist (Dennis

and Schnabel, 1996, Chapters 5 & 6), but they are hard to incorporate into the INLA routine. Hence, inference for such models is currently not possible within INLA.

Black box, convergence problems Although many INLA details are given in Rue *et al.* (2009), it is not possible for the end-user to understand all numerical details. On the one hand, it is an advantage of INLA that all approximations can easily be run by, *e.g.*, an epidemiologist using the provided `R` package as a black box. On the other hand, it is not always easy to detect the reason for an error message or a warning, if they occur.

For those who are interested in solving a specific `R-INLA` problem, two typical examples of error messages are shown in the following:

- `The Newton-Raphson optimizer did not converge.`
  This message reports problems when optimizing $\log \widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ to find $\boldsymbol{\theta}^*$, see Section 1.2. This issue can mostly be solved by setting reasonable `initial` values for the hyperparameters. Such values can be found by running the model using option `verbose = TRUE` and choosing the configuration of the hyperparameters in the last iteration as new `initial` values.
- `Matrix is not positive definite.`
  This problem reports that the numerical Hessian matrix computed for $\log \widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is not positive definite. Usually this issue is solved by increasing the step length h for this computation, which has a default of $h = 0.01$. For example, $h = 0.05$ can be chosen using the option `control.inla=list(h=0.05)`.

2 Modelling spatio-temporal surveillance data

In the following section we discuss hierarchical Bayesian models for spatial, temporal and spatio-temporal surveillance data, which are used in different contexts. The choice of an appropriate approach for a particular dataset depends on the aim of the analysis, the etiology of the disease and the spatial and temporal resolution of the available data. As already mentioned above, the fascinating fact is that all proposed models can be fitted within the INLA framework.

If we want to map endemic spatial and temporal patterns, disease mapping techniques as introduced in Section 2.1 can be used. Here, all variations in disease incidence are described by appropriate random effects. Hence, such models are also called parameter-driven (Cox, 1981). They are appropriate for the Swiss registry data, which are aggregated for 184 administrative regions and the Principality of Liechtenstein. These models were first established for purely spatial data.

For BT occurrence in Switzerland 2008 the exact location of each herd is known, see Section 2.2. Here, a two-dimensional location effect on a regular lattice is modelled directly in addition to the linear effects of individual, explanatory variables.

The models discussed in Section 2.3 are appropriate to analyze counts of infectious diseases, which usually show a regular pattern over time such as long-term trends or seasonality, but also occasional epidemic outbreaks. Here the disease mapping approach is not useful, as localized epidemics in the observed data cannot be modelled appropriately. The models originate from time series analysis and, hence, the purely temporal case is discussed first. For modelling epidemic outbreaks it might also be useful to apply so-called observation-driven models, where previous cases are included directly to model the infection mechanism (Cox, 1981).

2.1 Spatio-temporal disease mapping

We firstly introduce purely spatial disease mapping models and expand them to spatio-temporal data later.

Spatial disease mapping and ecological regression As a usual starting point for the analysis of aggregated data it is assumed that the number of cases y_i in region i is Poisson distributed with rate $E_i\lambda_i$

$$y_i \sim \text{Po}(E_i\lambda_i) \quad , i = 1, \ldots, I,$$

where E_i is the age-sex-adjusted expected number of cases in region i and λ_i is the relative risk (Knorr-Held and Becker, 2000). For the BVET data, only the number of herds at risk m_i is known and replaces E_i. It can easily be derived, that the ratio of y_i over m_i is the maximum likelihood estimate of λ_i. This ratio is called standard mortality ratio (SMR). Hence, a simple spatial analysis of surveillance data can be done by mapping these crude disease rates. However, the SMRs are difficult to interpret. Their expectation and variance is $\text{E}(\text{SMR}_i) = \text{E}(y_i)/m_i = \lambda_i$ and $\text{Var}(\text{SMR}_i) = \lambda_i/m_i$, respectively. This means that the variance of the SMRs grows for regions with a small population, so that the most extreme SMRs are typically found in low populated areas. Furthermore, the SMRs ignore spatial structure in the data.

A first approach to stabilize the estimation of the SMRs was made by Clayton and Kaldor (1987) (CK). They formulated a generalized linear mixed model with exchangeable random effects λ_i for each region, which are a priori assumed to be iid gamma distributed

$$\lambda_i \sim \text{Ga}(\tau_1, \tau_2).$$

The hyperparameters τ_1 and τ_2 are unknown, but can easily be estimated by empirical Bayes techniques (Clayton and Kaldor, 1987). This setting results in the posterior distribution

$$\hat{\lambda}_i | y \sim \text{Ga}(y_i + \hat{\tau}_1, m_i + \hat{\tau}_2)$$

with posterior mean

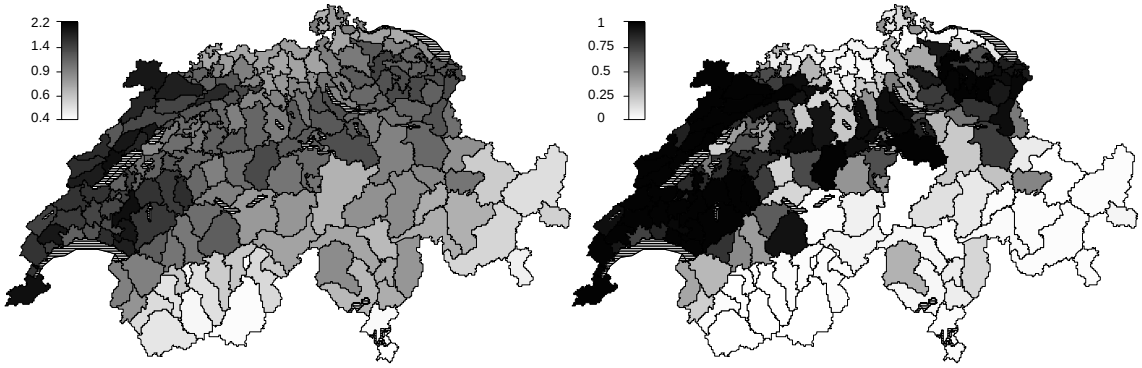
$$\hat{\lambda}_i = \frac{y_i + \hat{\tau}_1}{m_i + \hat{\tau}_2}$$

for the relative risk in region i. This estimator represents a weighted compromise between the SMRs and the overall mean rate. When rates are based upon a large y_i, the CK estimate differs little from the SMR. However, when there is little evidence that it differs from the overall mean, the estimate is drawn towards that value. Spatial autocorrelation is still not integrated in the risk estimates.

A fully Bayesian approach was firstly proposed by Besag *et al.* (1991) (BYM model). Here, the relative risk λ_i is modelled by a hierarchical Bayesian approach via a log link and and a linear predictor $\log(\lambda_i) = \eta_i$, see Section 1.1 (Clayton and Bernardinelli, 1992; Banerjee *et al.*, 2004). This linear predictor is usually specified as

$$\eta_i = \mu + \nu_i + \psi_i, \tag{7}$$

and follows a GMRF (Rue and Held, 2005; Held and Rue, 2010). The parameters $\nu_1, \ldots, \nu_I$ are iid normally distributed random effects with mean 0 and variance σ_ν^2. They account for uncorrelated variation in the respective region. On the contrary, the vector $\psi = (\psi_1, \ldots, \psi_I)^T$



(a) (b)

Figure 3: (a) Fitted spatial relative risk of BVD in Switzerland, 2008, and (b) the corresponding Bayesian p-values

has a joint prior

$$\pi(\boldsymbol{\psi}|\sigma_\psi^2) \propto \exp\left(-\frac{1}{2\sigma_\psi^2}\sum_{i\sim j}(\psi_i-\psi_j)^2\right), \tag{8}$$

where $i \sim j$ denotes spatial neighbours sharing a common boundary. The conditional distribution of ψ_i is

$$\psi_i|\psi_{-i},\sigma_\psi^2 \sim N\left(\frac{1}{n_i}\sum_{j:j\sim i}\psi_j, \frac{\sigma_\psi^2}{n_i}\right),$$

which shows that ψ_i "borrows strength" from its n_i spatial neighbours. The entries Q_{ij} of the precision matrix of $\boldsymbol{\psi}$ are -1, if $i \sim j$, and 0 otherwise. The diagonal elements are $Q_{ii} = n_i$. The vector $\boldsymbol{\psi}$ is an IGMRF, since all rows of $\boldsymbol{Q}$ sum to zero and the rank of $\boldsymbol{Q}$ is only $I-1$. The formulation (8) results in a spatial smoothing of the relative risk parameters where the degree of smoothing is determined through the parameter σ_ψ^2.

The usual choice of hyperpriors for the variance components σ_v^2 and σ_ψ^2 are highly dispersed priors like the inverse gamma distribution IGa(a,b). A discussion of different choices for a and b can be found in Bernardinelli *et al.* (1995a) and Paper IV.

The term $\exp(\nu_i+\psi_i)$ denotes the relative risk for a disease in region i. The relative risk for BVD in the year 2008 is shown in Figure 3a. Data on BVD were collected during an eradication program. Every cow in Switzerland was tested and, hence, the plot displays the true spatial dispersion of the disease risk. It ranges from 0.4 in the mountainous South and South-East of Switzerland to 2.2 in the North-West. For illustration, see the required `R-INLA` code for this example in Listing 1; for a more general introduction to the `R-INLA` code see Paper II.

Listing 1: `R-INLA` code for a BYM model

```
# Data: Y.BVD
# Offset: m.BVD
I <- length(Y.BVD)
region.nu <- seq(1,I,1)
region.psi <- seq(1,I,1)

# define components of the latent Gaussian field using f()
f.BVD <- Y.BVD~f(region.psi,model="besag",graph.file="switzerland.graph")+
               f(region.nu,model="iid")
```

```
# run function inla()
m.BVD <- inla(f.BVD,E=m.BVD,
         data=data.frame(Y.BVD,m.BVD,region.nu,region.psi),family="Poisson")
```

As an alternative to the estimated disease risk, the so-called Bayesian p-values can be mapped as a measure of the associated significance, see Figure 3b. They are defined as the posterior probabilities that the relative risk in a region i is larger than 1. Richardson *et al.* (2004) defined thresholds for such Bayesian p-values in order to decide, if the estimated relative risk for specific is significantly different from 1.

The standard BYM model can be extended in many ways. A joint analysis of several diseases was proposed in Held *et al.* (2005b) and Held *et al.* (2006). An extension of the model to a second, coarser spatial scale was studied by Langford *et al.* (1998) applying a multi-level approach. A similar strategy was used to assess the underreporting in BVD as to Swiss cantons in Paper V. A further, very important extension is the addition of area-specific covariates to the linear predictor (7) to explain the observed heterogeneity in the disease rates. This so-called ecological regression was introduced by Clayton *et al.* (1993). In epidemiological studies it is often used to study the association of disease risk with ecological covariates such as deprivation, urbanization or environmental factors. This concept was extended to nonparametric modelling of covariate effects by Natário and Knorr-Held (2003) and in Paper II. Note that the choice of appropriate hyperparameters is an important issue here.

A drawback concerning the analysis of a relationship between aggregated disease data and an area-level risk factor is the so-called ecological bias. It might happen that detected associations between risk factors and disease incidence on an aggregate level differ substantially from the (true) associations on an individual level. A well-known example are the suicide rates in Prussian provinces (1883–1890), which is illustrated in Morgenstern (1982).

Spatio-temporal disease mapping The most important extensions of the BYM model are those towards time, as it often happens that the temporal dynamics are even more pronounced than the spatial structure. A popular approach was made by Bernardinelli *et al.* (1995b), who deal with the analysis of the variation of insulin-dependent Diabetes Mellitus across Sardinia. A linear time trend β is added to the linear predictor (7) in addition to a random slope δ_i for each spatial unit:

$$\eta_{it} = \mu + \nu_i + \psi_i + (\beta + \delta_i) \cdot t,$$

where $i = 1, \ldots, I$ and $t = 1, \ldots, T$. Hence, there is a linear main time trend, but this trend can have a different slope for each region. Since this model is similar to a model with random intercept (ν_i or ψ_i) and random slope (δ_i), it is important to allow for a correlation between those two components (Hedeker and Gibbons, 2006, Section 4.4.2). Bernardinelli *et al.* (1995b) state that allowing for correlation will cause each estimate to be pulled towards the trend of areas with similar intercept. Depending on the assumption, whether the slope of the linear time trends varies smoothly in space or not, the δ_i's are correlated with the ψ_i's or ν_i's, respectively. This or similar models have been used in the literature by many authors. As an example see Assunção *et al.* (2001), who adopt an area-specific second-degree polynomial trend model for the evolution of disease rates over time to map and project the rates of visceral Leishmaniasis in Belo Horizonte. Applications of this model can also be found in Papers II, IV and V.

The assumption of a linear time trend is not plausible, if the data are observed over a long timespan. Hence, Knorr-Held (2000) proposed the incorporation of a nonparametric main time trend and space-time interactions for the analysis of lung cancer cases from Ohio (1968–

1988). The linear predictor has the general form

$$\eta_{it} = \mu + \nu_i + \psi_i + \gamma_t + \beta_t + \delta_{it}, \tag{9}$$

where the vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)^T$ has iid normally distributed components γ_t with mean 0 and variance σ_γ^2 for each timepoint $t = 1, \ldots, T$. The vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)^T$ is modelled by a so-called random walk of first order (RW1) (Rue and Held, 2005, Section 3.3.1). The respective prior can be written as

$$\pi(\boldsymbol{\beta}|\sigma_\beta^2) \propto \exp\left(-\frac{1}{2\sigma_\beta^2} \sum_{t=1}^{T-1} (\beta_{t+1} - \beta_t)^2 \right).$$

and forms an IGMRF. The conditional prior distribution of β_t at time $t+1$ depends only on the β's at time t and $t+2$ and the variance parameter σ_β^2.

Model (9) has an unstructured ($\boldsymbol{\gamma}$) and a structured ($\boldsymbol{\beta}$) temporal main effect in addition to the spatial main effects ($\boldsymbol{\psi}$ and $\boldsymbol{\nu}$). The specification of the spatio-temporal interaction term $\boldsymbol{\delta}$ depends on the assumption which spatial and temporal main effects are likely to interact. For example, if the disease incidence is different from region to region, but consistent over time, an interaction between $\boldsymbol{\nu}$ (iid) and $\boldsymbol{\beta}$ (RW1) can be assumed. In total four different types of interactions are possible. Their prior specification involving a Kronecker product formulation is described in detail in Knorr-Held (2000) and Paper IV.

This model formulation has proven to be useful in many different applications, *e.g.* for age-period-cohort models which deal with data stratified by age, year and region (Lagazio *et al.*, 2003; Schmid and Held, 2004). Lagazio *et al.* (2001) apply a variation of the model to lung cancer death certificate data in Tuscany, 1971–1994. Extensions of this model and a discussion of specific properties can be found in Papers IV and V.

2.2 GMRFs on a two-dimensional, regular lattice

If the exact location of an analysis unit in terms of eastings and northings is known in addition to individual covariates, we propose a two-dimensional GMRF on a regular lattice to model the spatial autocorrelation (Rue and Held, 2005, Section 3.4.2).

Originally, so-called Gaussian random field (GRF) priors have been proposed by Kammann and Wand (2003) for the analysis of point-referenced data to model a continuous spatial surface. However, Kneib and Fahrmeir (2006) state that GMRFs can be applied to point-referenced data by a discretization of the observation area. They can be defined on a regular or irregular lattice and guarantee a more parsimonious and computationally less expensive modelling than the GRFs due to the additional Markov property, see Section 1.1. In the regular case, this means that the known point locations have to be assigned to a (dense) two-dimensional lattice with A rows and B columns. Similar models are often used in, *e.g.*, image analysis, where nodes represent pixels (Besag and Kooperberg, 1995). As an example a map of the locations of all Swiss cattle farms on a regular 100×80 lattice is shown in Figure 4.

Priors with different complexity can be imposed on this two-dimensional lattice. The simplest option is to use a linear spatial trend, where the response of a unit depends on the average response of the units in the same row and the same column. For a more localized spatial effect, an intrinsic second order random walk prior (RW2d) can be used (Rue and Held, 2005, Section 3.4.2). Here, the full conditional mean of one node on the lattice depends on the four direct neighbours on the lattice, the four diagonal neighbours on the lattice and the

Figure 4: Locations of Swiss cattle farms on a two-dimensional, regular lattice

four direct neighbours of second order. Hence, a node borrows strength from its 12 closest spatial neighbours. This model is useful for a local, slowly varying spatial surface (Besag and Kooperberg, 1995). The size of the spatial effect is controlled by a smoothing variance.

In Paper III this methodology is applied to BT cases in Switzerland in 2008. The location of each Swiss farm is known in addition to the information, whether a BT case was detected. Vaccination coverage, surveillance intensity and altitude are considered as potential explanatory variables. For illustration, see in Listing 2, how an RW2d can be specified for data with exact spatial location.

Listing 2: `R-INLA` code for model `rw2d`

```
# construct discrete lattice of farm locations from longitude/latitude data
loc.farms <- data.frame(longitude,latitude)

# define dimensions of lattice
# beware of internal INLA lattice: nrow -> max(longitude), ncol -> max(latitude)
ncol <- 80   # choice is due to the shape of Switzerland
nrow <- 100
dim.long <- range(loc.farms[,1])
dim.lat <- range(loc.farms[,2])
cutpoints.long <- seq((dim.long[1]-0.1),(dim.long[2]+0.1),length.out=nrow)
cutpoints.lat <- seq((dim.lat[1]-0.1),(dim.lat[2]+0.1),length.out=ncol)

# assign farm locations to discrete lattice
lattice.long <- as.numeric(cut(loc.farms[,1],cutpoints.long))
lattice.lat <- as.numeric(cut(loc.farms[,2],cutpoints.lat))

# transform lattice into one-dimensional node for INLA
nodes <- numeric(dim(loc.farms)[1])

for(i in 1:length(loc.farms[,1])){
nodes[i] <- inla.lattice2node(lattice.long[i],lattice.lat[i],nrow=nrow,
                                                              ncol=ncol)}

# Data: Y.BT
# define random walk of second order "rw2d"
f.BT <- Y.BT~f(nodes, model = "rw2d", nrow=nrow, ncol=ncol)
```

2.3 Modelling approaches for infectious disease counts

In the following we discuss parameter- and observation-driven models for time series of infectious disease counts and discuss their generalization to multivariate data.

Parameter-driven models It is possible to model a time series of infectious disease counts y_t ($t = 1, \ldots, T$) by a purely parameter-driven model assuming a Poisson distribution $y_t \sim \text{Po}(\exp(\eta_t))$ for the observed counts (Zeger, 1988). Here it is assumed that an unobserved autocorrelated mechanism drives the infection process. The basic model can be formulated using two stages

$$\begin{array}{ll} \text{Stage 1:} & \eta_t = \alpha + \zeta_t \\ \text{Stage 2:} & \zeta_t = \kappa \cdot \zeta_{t-1} + \epsilon_t, \end{array} \tag{10}$$

where $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_T)^T$ forms an autoregressive process of first order (AR1). The errors $\epsilon_1, \ldots, \epsilon_T$ are assumed to be iid normally distributed. Nelson and Leroux (2006) discuss maximum likelihood inference in such generalized linear mixed models and compare the parameter estimation of various methods for the incidence of Polio in the USA during 1970–1983. Models with a hidden stochastic process, as (10), can also be understood as generalized dynamic models (Fahrmeir, 1992). In a dynamic model, both parameter variation and available data information are described in a probabilistic way. The formulation of such models in a hierarchical Bayesian framework with latent Gaussian field is very convenient (Gamerman, 1998; Hay and Pettitt, 2001).

Consider a monthly time series of Salmonella Agona (SA) cases in the UK, 1990–1995. The lines of `R-INLA` code in Listing 3 show, how model (10) can easily be fitted using INLA.

Listing 3: `R-INLA` code for a generalized dynamic model

```
# salmonella data are stored in R package surveillance
library(surveillance)
data(salmonella.agona)

Y.SA <- salmonella.agona$observed
ar1.zeta <- 1:length(Y.SA)

f.SA <- Y.SA~f(ar1.zeta,model="ar1")
m.SA <- inla(f.SA,family="Poisson",data=data.frame(Y.SA,ar1.zeta))
```

Observation-driven models As an alternative, so-called mechanistic, compartmental models are used, which concentrate on the temporal disease dynamics (Cliff and Haggett, 1988; Grenfell *et al.*, 1995). In the famous SIR models the compartments are those individuals susceptible to infection (S), those currently infectious (I), and those recovered and immune to infection (R) (Anderson and May, 1991). These models describe directly the spread from person-to-person on an individual level. However, a major requirement is that the epidemic process is completely observed. One has to know the number of infected and the number of susceptible individuals at each timepoint. As a solution such models can be approximated by a branching process, where an unlimited amount of susceptibles is assumed.

Held *et al.* (2005a) proposed a model based on a branching process with immigration. To describe local epidemics previous counts enter additively into the linear predictor. Hence, the infection process is governed by an observation-driven mechanism. The basic assumption for a time series of counts is that the counts are Poisson or negative binomially distributed with

mean

$$\mu_t = \kappa \cdot y_{t-1} + \exp(\eta_t). \tag{11}$$

This model is a mixture of epidemic ($\kappa \cdot y_{t-1}$) and endemic ($\exp(\eta_t)$) components. The time series is stationary for $|\kappa| < 1$. Given this condition, the endemic incidence is persistent with a stable temporal, perhaps seasonal pattern. The epidemic incidence will break out occasionally. Classical maximum likelihood inference is possible in this setting.

Inclusion of network data Often surveillance data are multivariate, as time series of counts are available for a number of geographic regions. It is known that networks of moving individuals between these regions represent a potential risk of disease transmission (Fèvre *et al.*, 2006). As an example, many authors have assessed cattle trade networks as a source of spatio-temporal disease spread for, *e.g.*, Bovine Tuberculosis and Foot-and-Mouth disease (Ferguson *et al.*, 2001; Gilbert *et al.*, 2005; Jewell *et al.*, 2009). Mostly the disease dynamics in the network are examined using stochastic simulation models. However, we are interested in spatio-temporal statistical models that can directly relate network information to the observed disease data.

Both models (10) and (11) can be extended to the multivariate case including network data. For (11), this is discussed and applied to real data examples on, *e.g.*, Influenza and Meningococcal disease in Paul *et al.* (2008) and Paul and Held (2011). In Paper VI the purely parameter-driven approach (10) is expanded to the multivariate case using vector-autoregressive models as a building block (Lütkepohl, 2005). The impact of cattle trade on the spatio-temporal spread of an infectious disease among Swiss cows is assessed.

3 Model choice

The formulation of criteria for model choice in Bayesian hierarchical models is an important issue. We review the well-known deviance information criterion (DIC) in Section 3.1 and various so-called predictive measures in Section 3.2 and discuss their derivation by INLA in Section 3.3.

3.1 Deviance information criterion

The DIC is a Bayesian measure of model fit and complexity. It was proposed by Spiegelhalter *et al.* (2002) and is heavily used for model choice in epidemiological applications (Knorr-Held, 2000; Lagazio *et al.*, 2001; Held *et al.*, 2005b). It is defined as the sum of a measure of model fit, the posterior mean of the deviance $\overline{D}$, and a measure of model complexity, the effective number of parameters p_D. The lower the DIC, the better is the model. The quantity p_D is used in models like the complex hierarchical models, where the absolute number of parameters is not as clearly defined as in the normal linear model, for example. A smaller mean deviance indicates a better model fit, but it decreases with an increasing number of parameters. Hence, the DIC provides a trade-off between model fit and complexity. However, the DIC has been criticized to underpenalize complex models (Plummer, 2008).

The DIC can be obtained from `R-INLA` using the option `control.compute=list(dic=TRUE)`. Its calculation in INLA is described in Rue *et al.* (2009, Section 6.4).

3.2 Evaluation of predictive performance

Proper scoring rules Another approach towards model choice is the use of scores to evaluate probabilistic forecasts obtained from the model (Gneiting and Raftery, 2007; Gneiting *et al.*, 2007). They assign a numerical score based on the predictive distribution and the actually observed count to each prediction. Two important properties, which are addressed simultaneously, are sharpness and calibration (Gneiting *et al.*, 2007). Sharpness is a measure for the concentration of the predictive distribution. Calibration assesses the statistical consistency between the predictive distribution and the observed count. A scoring rule is called proper, if its expected value under the predictive distribution becomes minimal, if the observed count is indeed a realization from the predictive distribution (Gneiting and Raftery, 2007). An attractive feature of the scores is that they can be applied to parametric and non-parametric settings and do not require the models to be nested or related (Czado *et al.*, 2009). Frequently used examples of scoring rules are the squared error score (SES), the logarithmic score (logS), the Dawid-Sebastiani score (DSS) and the ranked probability score (RPS). Czado *et al.* (2009) derived their formulas for count data:

$$\text{SES}(P, y) = (y - \mu_P)^2$$

$$\text{logS}(P, y) = -\log(P(Y = y))$$

$$\text{RPS}(P, y) = \sum_{k=0}^{\infty} (P(Y \leq k) - \mathbf{1}(y \leq k))^2$$

$$\text{DSS}(P, y) = \tfrac{1}{2} \cdot (\log(\sigma_P^2) + ((y - \mu_P)/\sigma_P)^2).$$

Here, P is the predictive probability distribution, μ_P and σ_P are its mean and standard deviation and y is the observed count. The different strengths and weaknesses of these scores are discussed in Gneiting and Raftery (2007).

Note that scores can be used in the context of leave-one-out cross-validation for, *e.g.*, disease mapping models as in Section 2.1. Here a cross-validated mean score is used for model choice. However, they are also useful to assess one-step-ahead predictions for a time series of counts as introduced in Section 2.3 by computing a mean score for successive one-step-ahead forecasts. For the logarithmic score there are interesting relationships to the Akaike (AIC) and the Bayesian information criterion (BIC). These criteria are frequently used for model choice within a maximum likelihood framework (Fahrmeir and Tutz, 2001). According to Stone (1977), the cross-validated mean logarithmic score is asymptotically equivalent to the AIC, if the observations are independent. Furthermore, by re-formulation of a joint density into a product of one-step-ahead predictive densities, Dawid (1984) showed that the mean of the one-step-ahead log scores is asymptotically equivalent to the BIC (see also Section 7.1 in Gneiting and Raftery, 2007).

Tools to assess calibration Some measures proposed in the literature assess only the calibration of the predictive distribution. Gneiting *et al.* (2007) discuss the probability integral transform (PIT), which is the value of the predictive cumulative distribution at the observed count $P(Y \leq y)$. They state that if the actual observations were drawn from their predictive distributions, the PIT values have a standard uniform distribution. As a diagnostic tool, a histogram of empirically obtained leave-one-out cross-validated PIT values can be plotted and checked for uniformity. Deviations from uniformity indicate under- or overdispersed predictive distributions. Czado *et al.* (2009) propose a version of the PIT histogram that is tailored to

count data.
Held *et al.* (2011) propose a significance test based on scoring rules to assess the calibration of continuous predictive distributions. For count data it is possible to assess the calibration by a score regression approach based on the DSS: Held *et al.* (2011) state that if μ_P and σ_P match the corresponding quantities of the true data-generating distribution, the expectation of the DSS depends only on $\log(\sigma_P)$. This relation can be checked by a linear regression.

3.3 Computation of proper scoring rules with INLA

In INLA, the leave-one-out cross-validated quantities $P(Y = y)$ (`cpo`) and $P(Y \leq y)$ (`pit`) can be computed directly without re-running the model on reduced data. Numerical details are described in Rue *et al.* (2009, Section 6.3) and Paper I. In practice, they are obtained using the option `control.compute=list(cpo=TRUE)`.
For one-step-ahead forecasts the predictive distribution can be obtained by coding the respective observation as `NA`; more details on the `R-INLA` code for the computation of one-step-ahead scores are given below. In the following the term y_t denotes the truly observed count at time t of a time series and $\boldsymbol{y} = (y_1, \ldots, y_T)^T$ the vector of all observed counts. The vector $\boldsymbol{y}_{-T}$ contains all observations up to time $T-1$, *i.e.* the history at time T.

Calculating the predictive density $P(y_T|\boldsymbol{y}_{-T})$ from INLA approximations of $\pi(\lambda_T|\boldsymbol{y}_{-T})$ Suppose, $y_t \sim \text{Po}(\lambda_t = \exp(\eta_t))$. We use the output obtained from a model fitted with data $y_1, \ldots, y_T$, where the last observation y_T was coded as missing with `NA`.
The predictive distribution of y_T can always be written as

$$P(y_T|\boldsymbol{y}_{-T}) = \int \pi(y_T|\lambda_T, \boldsymbol{y}_{-T})\ \pi(\lambda_T|\boldsymbol{y}_{-T})\ d\lambda_T.$$

A discretized approximation of the marginal of the linear predictor $\pi(\eta_T|\boldsymbol{y}_{-T})$ at J nodes is obtained from INLA, when setting the option `control.predictor=list(compute=TRUE)`. This marginal can be transformed into a marginal of $\lambda_T = \exp(\eta_T)$ using the function `inla.tmarginal()`. Using this output, we can approximate $P(y_T|\boldsymbol{y}_{-T})$ by a finite sum

$$P(y_T|\boldsymbol{y}_{-T}) \approx \sum_{j=1}^{J} \pi(y_T|\lambda_T^{(j)}, \boldsymbol{y}_{-T})\ \pi(\lambda_T^{(j)}|\boldsymbol{y}_{-T})\ \Delta_j. \tag{12}$$

Calculating μ_P and σ_P^2 from INLA approximations of $\pi(\lambda_T|\boldsymbol{y}_{-T})$ Using the function `inla.emarginal()` we can obtain the mean and standard deviation of $\pi(\lambda_T|\boldsymbol{y}_{-T})$. Applying the law of iterated expectations (*e.g.* Billingsley, 1986), we can then calculate μ_P as

$$\mu_P = \text{E}(y_T|\boldsymbol{y}_{-T}) = \text{E}(\text{E}(y_T|\lambda_T, \boldsymbol{y}_{-T})) = \text{E}(\lambda_T|\boldsymbol{y}_{-T}). \tag{13}$$

The predictive variance σ_P^2 can be derived using the law of total variance

$$\text{Var}(y_T|\boldsymbol{y}_{-T}) = \text{E}(\text{Var}(y_T|\lambda_T, \boldsymbol{y}_{-T})) + \text{Var}(\text{E}(y_T|\lambda_T, \boldsymbol{y}_{-T})) = \text{E}(\lambda_T|\boldsymbol{y}_{-T}) + \text{Var}(\lambda_T|\boldsymbol{y}_{-T}). \tag{14}$$

`R-INLA` code for the calculation of one-step-ahead scores Consider the time series of Salmonella Agona (SA) cases in the UK from Section 2.3 as an example. The `R-INLA` code in Listing 4

shows, how a single one-step-ahead prediction for the last timepoint T is made and the SES, logS, DSS and RPS can be calculated.

Listing 4: R-INLA code for one-step-ahead scores

```
## make a one-step-ahead prediction for the last timepoint T
T <- length(Y.SA)

# code last observation manually as NA
Y.SA.NA <- c(Y.SA[1:(T-1)],NA)

# run model with NA data
f.SA.NA <- Y.SA.NA~1+f(ar1.zeta,model="ar1")
m.SA.NA <- inla(f.SA.NA,family="Poisson",data=data.frame(Y.SA.NA,ar1.zeta),
                                              control.predictor=list(compute=1))

# obtain marginal of T via transformation
marg.T <- inla.tmarginal(function(x)exp(x),m.SA.NA$marginals.fitted.values[[T]])

# calculate mean of predictive distribution using formula (13)
mu_P <- mu_lambda <- inla.emarginal(function(x) x, marg.T)

# calculate standard deviation of predictive distribution using formula (14)
var_lambda <- inla.emarginal(function(x) x^2, marg.T)-mu_P^2
sd_P <- sqrt(mu_P + var_lambda)

## calculate SES
SES <- (Y.SA[T]-mu_P)^2

## calculate logS
# approximate sum (12)
f <- 0
for (j in 1:(length(marg.T[,1])-1)){
f <- f+dpois(Y.SA[T],(marg.T[j,1]+marg.T[j+1,1])/2)*trapz(marg.T[j:(j+1),1],
                                                          marg.T[j:(j+1),2])}
logS <- -log(f)

## calculate DSS
DSS <- 1/2*(log(sd_P^2)+((Y.SA[T]-mu_P)/sd_P)^2)

## calculate RPS
# initialize vector for P(Y=k)
p.k <- 0
# run loop over all k
k <- 0
# stop, if probability becomes very small
f <- 1
while (round(f,digits=40)> 0){
f <- 0
# calculate P(Y=k) using formula (12)
for (j in 1:(length(marg.T[,1])-1)){
f <- f+dpois(k,(marg.T[j,1]+marg.T[j+1,1])/2)*trapz(marg.T[j:(j+1),1],
                                                    marg.T[j:(j+1),2])}
p.k <- c(p.k,f)
k<-k+1}
p.k<-p.k[-1]

# evaluate RPS
RPS <- 0
for(k in 0:(length(p.k)-1)){
RPS <- RPS+(sum(p.k[1:(k+1)])-(as.numeric(Y.SA[T]<=k)))^2}
```

As already mentioned above, it is common to perform several one-step-ahead predictions and to base the model comparison on the resulting mean scores.

Thesis Summary

This thesis consists of six papers. Their content and contribution are briefly summarized below. If not stated otherwise, approximate Bayesian inference is implemented using INLA.

Paper I

Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA by L. Held, B. Schrödle and H. Rue.
This paper discusses criticism and comparison of Bayesian hierarchical models and shows that leave-one-out cross-validatory predictive checks are superior to so-called posterior predictive checks, where all observations are used. Furthermore, it is shown how leave-one-out predictive quantities are directly approximated by INLA without re-running the model. Additionally, elaborate re-sampling techniques proposed in the MCMC framework to avoid manual crossvalidation are reviewed. The results of INLA and various MCMC approaches are compared for a BYM model applied to BVD cases in Switzerland, 2008.

This article is published in a Festschrift in the honour of Prof. Dr. L. Fahrmeir. The contribution was initiated by L. Held and he wrote the more technical parts of the draft. All code for the MCMC and INLA analyses was implemented and evaluated by me. We finalized the manuscript together. H. Rue gave us input on the INLA approximations concerning leave-one-out cross-validatory predictive quantities.

Paper II

A primer on disease mapping and ecological regression using INLA by B. Schrödle and L. Held.
This paper introduces models for spatial and spatio-temporal disease mapping, which are frequently used by epidemiologists. To show the high usability of INLA, the `R-INLA` code for a case study on Coxiellosis among Swiss cows is displayed. Furthermore, it is discussed how an explanatory variable can be comprised in a linear or nonparametric fashion. To facilitate model choice, the usage of various criteria available directly from INLA is shown. Additionally, the computer time and accuracy of different INLA options are examined. Supplementary material containing all `R-INLA` code is provided.

This paper is based on two talks, which L. Held and me gave at the workshop “Statistical Computing" in Reisensburg, Germany, in July 2009 and is published in a special issue at this occasion. Supervised by L. Held I did all the writing.

Paper III

Spatial analysis of Bluetongue cases and vaccination of Swiss cattle in 2008 and 2009 by K. Willgert, B. Schrödle and H. Schwermer.
This paper deals with the association between vaccination coverage and BT occurrence on Swiss farms in 2008 and 2009. We regress the point-referenced BT cases on the vaccination coverage of each farm. An individual surveillance indicator and the altitude of each

holding are additionally included into the model as potential confounders. Furthermore, a two-dimensional location effect on a regular lattice is brought in to adjust for spatial autocorrelation. As priors we propose to use a linear trend or second order random walk on a regular lattice. A similar analysis is conducted for 17 so-called BT regions, which are supposed to be homogeneous with respect to vaccination coverage and surveillance efforts. The results suggest that a higher vaccination coverage reduces the risk of a BT infection.

This paper is based on the master thesis of K. Willgert (Royal Veterinary College, London) which was elaborated in cooperation with the Swiss Federal Veterinary Office (BVET). The BT data were extracted and pre-processed by K. Willgert. The methods for a statistical analysis of the data were proposed and carried out by me. The draft was written by K. Willgert, except for the sections on statistical methods and the description of the results. The manuscript was finalized by K. Willgert, H. Schwermer and me. H. Schwermer supported us with his knowledge on veterinary epidemiology.

Paper IV

Spatio-temporal disease mapping using INLA by B. Schrödle and L. Held.
In this paper we discuss various choices of the main time effect (linear, RW1, RW2) within spatio-temporal disease mapping models and compare their properties. A specific emphasis is on the form of the spatio-temporal interaction term in nonparametric models and its INLA implementation requiring user-defined structure matrices and linear constraints. The methodology is applied to cases of Salmonellosis in Swiss cows during 1991–2008. Model choice is performed using the DIC.

This paper is based on a talk of L. Held at the “TIES" conference in Bologna, Italy, in July 2009 and is published in a special issue at this occasion. With support by L. Held, I conducted all analyses and most of the writing.

Paper V

Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: A case-study by B. Schrödle, L. Held, A. Riebler and J. Danuser
This paper deals with underreporting in the BVD cases, 2003–2007, with regard to the affiliation of a Swiss region to a canton. As Switzerland is a confederation of 26 cantons which have their own veterinary authorities, it is supposed that there are differences in the realization of federal veterinary legislation. State-of-the-art spatio-temporal disease mapping models for administrative regions are extended to a coarser, cantonal grid using a multi-level modelling approach. Reasonable choices regarding the properties of the spatio-temporal interaction term are discussed. Purely regional models are compared to multilevel models by DIC, logarithmic score and PIT histograms. The results show that underreporting is present in several Swiss cantons. Furthermore, a comparison of INLA approximations and MCMC histograms with regard to the accuracy of the parameter estimates and the usability of both approaches in practice is conducted.

In this paper the ideas concerning underreporting in the BVD data were developed by me. L. Held gave much input with regard to the statistical methodology. The MCMC algorithm for the nonparametric space-time model was implemented by A. Riebler and the results were evaluated together. The manuscript was finalized by me in agreement with J. Danuser from the BVET, who gave insight into the veterinary legislation in Switzerland.

Paper VI

Assessing the impact of network data on the spatio-temporal spread of infectious diseases
by B. Schrödle, L. Held and H. Rue.
We propose parameter- and observation-driven models for multivariate infectious disease counts, where information on networks between regions can be directly included. A specific emphasis is on the formulation of parameter-driven hierarchical models based on vector-autoregressive processes. Model comparison is facilitated by assessing the one-step-ahead predictive performance of forecasts using proper scoring rules and a test of calibration based on the DSS. Using this methodology it is shown that a spatio-temporal spread of Coxiellosis among Swiss cows via cattle trade takes place. Furthermore, it turns out that the predictive performance of parameter-driven models is superior to so-called observation-driven models, where previous counts are directly included in the model. Supplementary material on the implementation of parameter-driven models in INLA is provided.

The idea for this paper was developed during a research visit at the group of H. Rue at the Norwegian University of Science and Technology in Trondheim. H. Rue gave much input concerning the implementation of the models in INLA. I implemented the methods and wrote a draft of the paper. The paper was finalized jointly with L. Held.

An earlier version of this manuscript is published in the Proceedings of the “25th International Workshop on Statistical Modelling" in Glasgow in July 2010. It can be found in the Appendix of this dissertation.

References

Anderson, R. M. and May, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford.

Assunção, R. M., Reis, I. A. and Oliveira, C. (2001). Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model, *Statistics in Medicine* **20**(15): 2319–2335.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, London.

Bernardinelli, L., Clayton, D. and Montomoli, C. (1995a). Bayesian estimates of disease maps: how important are priors?, *Statistics in Medicine* **14**: 2411–2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. and Ghislandi, M. (1995b). Bayesian analysis of space-time variation in disease risk, *Statistics in Medicine* **14**: 2433–2443.

Bernet, D. (2005). Salmonellen-Untersuchungen bei Rindern in den Jahren 2000 bis 2004. Auswertung der gemeldeten Untersuchungsergebnisse in der zentralen Datenbank für Labordiagnostik (in German), *BVET internal paper*.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregression, *Biometrika* **82**(4): 733–746.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics* **43**(1): 1–59.

Billingsley, P. (1986). *Probability and Measure*, John Wiley & Sons, New York.

Buehler, J. W. (2008). Surveillance, *in* K. J. Rothman, S. Greenland and T. L. Lash (eds), *Modern Epidemiology*, Lippincott Williams & Wilkins, Philadelphia.

BVET (2006). BVD im Detail. Available from: `http://www.bvet.admin.ch/gesundheit_tiere/00286/00288`.

Clayton, D. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk, *in* J. Cuzick et al. (eds), *Geographical and Environmental Epidemiology. Methods for Small Area Studies*, Oxford University Press, pp. 205–220.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics* **43**: 671–681.

Clayton, D., Bernardinelli, L. and Montomoli, C. (1993). Spatial correlation in ecological analysis, *International Journal of Epidemiology* **22**(6): 1193–1202.

Cliff, A. D. and Haggett, P. (1988). *Atlas of Disease Distributions*, Blackwell, Oxford.

Cohen-Sabas, C. H., Heim, D., Zurbriggen, A. and Stärk, K. D. C. (2004). Age-period-cohort analysis of the Bovine Spongiform Encephalopathy BSE epidemic in Switzerland, *Preventive Veterinary Medicine* **66**: 19–33.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments (with discussion and reply), *Scandinavian Journal of Statistics* **8**(2): 93–115.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data, *Biometrics* **65**(4): 1254–1261.

Dawid, A. P. (1984). Statistical theory: the prequential approach, *Journal of the Royal Statistical Society - Series A* **147**: 278–292.

Dennis, J. E. and Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Society for Industrial and Applied Mathematics, Philadelphia.

Doherr, M. G. and Audige, L. (2001). Monitoring and surveillance for rare health-related events: a review from the veterinary perspective, *Philosophical Transactions of the Royal Society B* **356**: 1097–1106.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models, *Journal of the American Statistical Association* **87**(418): 501–509.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer, Berlin.

Ferguson, N. M., Donnelly, C. A. and Anderson, R. M. (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions, *Science* **292**: 1155–1160.

Fèvre, E. M., de C. Bronsvoort, B. M., Hamilton, K. A. and Cleaveland, S. (2006). Animal movements and the spread of infectious diseases, *Trends in Microbiology* **14**(3): 125–131.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state-space modelling, *Biometrika* **93**: 827–841.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models, *Biometrika* **85**(1): 215–227.

Gelman, A. E., Carlin, J. P., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton.

Giesecke, J. (2002). *Modern Infectious Disease Epidemiology*, Hodder Arnold, London.

Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R. and Wint, W. (2005). Cattle movement and bovine tuberculosis in Great Britain, *Nature* **435**: 491–496.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**(477): 359–378.

Gneiting, T., Balabdoui, F. and Raftery, E. (2007). Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society - Series B* **69**(2): 243–268.

Grenfell, B. T., Kleczkowski, A., Giligan, C. A. and Bolker, B. M. (1995). Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases, *Statistical Methods in Medical Research* **4**: 160–183.

Hay, J. L. and Pettitt, A. N. (2001). Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease, *Biostatistics* **2**(4): 433–444.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*, Wiley, New Jersey.

Held, L. and Rue, H. (2010). Conditional and intrinsic autoregressions, *in* A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp (eds), *Handbook of Spatial Statistics*, Chapman & Hall/CRC, Boca Raton.

Held, L., Graziano, G., Frank, C. and Rue, H. (2006). Joint spatial analysis of gastrointestinal infectious diseases, *Statistical Methods in Medical Research* **15**(5): 465–480.

Held, L., Höhle, M. and Hofmann, M. (2005a). A statistical framework for the analysis of multivariate infectious disease surveillance counts, *Statistical Modelling* **5**: 187–199.

Held, L., Natário, I., Fenton, S. E., Rue, H. and Becker, N. (2005b). Towards joint disease mapping, *Statistical Methods in Medical Research* **14**: 61–82.

Held, L., Rufibach, K. and Balabdaoui, F. (2011). A score regression approach to assess calibration of continuous probabilistic predictions, *Biometrics* **66**: 1295–1305.

Jewell, C. P., Kypraios, T., Neal, P. and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases, *Bayesian Analysis* **4**: 465–496.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models, *Journal of the Royal Statistical Society - Series C* **52**: 1–18.

Keeling, M. J. and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, Princeton and Oxford.

Kneib, T. and Fahrmeir, L. (2006). Structured additive regression for categorical space-time data: a mixed model approach, *Biometrics* **62**: 109–118.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine* **19**: 2555–2567.

Knorr-Held, L. and Becker, N. (2000). Bayesian modelling of spatial heterogeneity in disease maps with application to German cancer mortality data, *Allgemeines Statistisches Archiv* **84**: 121–140.

Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping, *Scandinavian Journal of Statistics* **29**(4): 597–614.

Lagazio, C., Biggeri, A. and Dreassi, E. (2003). Age-period-cohort models and disease mapping, *Environmetrics* **14**: 475–490.

Lagazio, C., Dreassi, E. and Biggeri, A. (2001). A hierarchical Bayesian model for space-time variation of disease risk, *Statistical Modelling* **1**: 17–29.

Langford, I. H., Bentham, G. and McDonald, A.-L. (1998). Multi-level modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European community, *Statistics in Medicine* **17**: 41–57.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

Morgenstern, H. (1982). Uses of ecologic analysis in epidemiologic research, *American Journal of Public Health* **72**(12): 1336–1344.

Natário, I. and Knorr-Held, L. (2003). Non-parametric ecological regression and spatial variation, *Biometrical Journal* **45**(6): 670–688.

Nelson, K. P. and Leroux, B. G. (2006). Statistical models for autocorrelated data, *Statistics in Medicine* **25**: 1413–1430.

Nobre, A. A., Schmidt, A. M. and Lopes, H. F. (2005). Spatio-temporal models for mapping the incidence of Malaria in Pará, *Environmetrics* **16**: 291–304.

Paul, M. and Held, L. (2011). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts, *Statistics in Medicine* **30**(10): 1118–1136.

Paul, M., Held, L. and Toschke, A. M. (2008). Multivariate modelling of infectious disease surveillance data, *Statistics in Medicine* **27**: 6250–6267.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison, *Biostatistics* **9**(3): 523–539.

Presi, P. and Heim, D. (2010). BVD eradication in Switzerland - a new approach, *Veterinary Microbiology* **142**(1): 137–142.

R Development Core Team (2005). *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: `http://www.R-project.org`.

Raoult, D., Marrie, T. J. and Mege, J. L. (2005). Natural history and pathophysiology of Q fever, *The Lancet Infectious Diseases* **5**(4): 219–226.

Richardson, S., Thomson, A., Best, N. and Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies, *Environmental Health Perspectives* **112**: 1016–1025.

Riebler, A., Held, L. and Rue, H. (2010). Correlated multivariate age-period-cohort models, *Technical report*, University of Zurich, Switzerland.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields, *Journal of the Royal Statistical Society - Series B* **63**(2): 325–338.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*, Chapman & Hall/CRC, London.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models, *Journal of Statistical Planning and Inference* **137**: 3177–3192.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society - Series B* **71**: 319–392.

Salman, M. D., Stärk, K. D. C. and Zepeda, C. (2003). Quality assurance applied to animal disease surveillance systems, *Revue scientifique et technique - Office international des épizooties* **22**: 689–696.

Schmid, V. (2011). Voxel based adaptive spatio-temporal modelling of perfusion cardiovascular MRI, *IEEE Transaction on Medical Imaging*. DOI: 10.1109/TMI.2011.2109733.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data, *Biometrics* **60**: 1034–1042.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society - Series B* **64**(4): 583–639.

Stärk, K. D. C. (1996). Animal health monitoring and surveillance in Switzerland, *Aust. Vet. J* **73**: 96–97.

Steinsland, I. and Jensen, H. (2010). Utilizing Gaussian Markov random field properties of Bayesian animal models, *Biometrics* **66**: 763–771.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and akaike's criterion, *J. R. Statist. Soc. B* **39**(1): 44–47.

Stroup, D. F., Brookmeyer, R. and Kalsbeel, W. D. (2004). Public health surveillance in action: a framework, *in* R. Brookmeyer and D. Stroup (eds), *Monitoring the health of populations*, Oxford University Press.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**(393): 82–86.

Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generlized models, *Biometrika* **63**(1): 27–32.

Wegener, H. C., Hald, T., Wong, D. L. F., Madsen, M., Korsgaard, H., Bager, F., Gerner-Smidt, P. and Mølbak, K. (2003). *Salmonella* control programs in Denmark, *Emerging Infectious Diseases* **9**: 774–780.

Zeger, S. L. (1988). A regression model for time series counts, *Biometrika* **75**: 621–629.

Zimmerli, U., Presi, P. and Heim, D. (2010). BVD eradication campaign in Switzerland: first results and outlook, *Schweizer Archiv für Tierheilkunde* **151**(1): 5–11.

PAPER I

Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA

Leonhard Held, Birgit Schrödle & Håvard Rue

Pre-print version of a paper published in *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir, Thomas Kneib & Gerhard Tutz (eds.)*, 2010, Physica-Verlag, Heidelberg, ISBN 978-3-7908-2412-4.

Posterior and Cross-validatory Predictive Checks: A Comparison of MCMC and INLA

Leonhard Held[1], Birgit Schrödle[1] and Håvard Rue[2]

[1] Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Switzerland

[2] Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim

Model criticism and comparison of Bayesian hierarchical models is often based on posterior or leave-one-out cross-validatory predictive checks. Cross-validatory checks are usually preferred because posterior predictive checks are difficult to assess and tend to be too conservative. However, techniques for statistical inference in such models often try to avoid full (manual) leave-one-out cross-validation, since it is very time-consuming. In this paper we will compare two approaches for estimating Bayesian hierarchical models: Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximations (INLA). We review how both approaches allow for the computation of leave-one-out cross-validatory checks without re-running the model for each observation in turn. We then empirically compare the two approaches in an extensive case study analysing the spatial distribution of bovine viral diarrhoe (BVD) among cows in Switzerland.

Keywords: Bayesian hierarchical models; INLA; Leave-one-out cross-validation; MCMC; Posterior predictive model checks

1 Introduction

Bayesian hierarchical models are widely used in applied statistics. Inference is typically based on Markov chain Monte Carlo (MCMC), a computer-intensive simulation-based approach. However, integrated nested Laplace approximations (INLA) are a promising alternative to inference via MCMC in latent Gaussian models (Rue *et al.*, 2009). The methodology is particularly attractive if the latent Gaussian model is a Gaussian Markov random field (GMRF) (Rue and Held, 2005). In contrast to empirical Bayes approaches (Fahrmeir *et al.*, 2004), the INLA approach incorporates posterior uncertainty with respect to hyperparameters. Examples where INLA is applicable include generalized linear mixed models (Breslow and Clayton, 1993), disease mapping (Besag *et al.*, 1991) including ecological regression (Clayton and Bernardinelli, 1992; Natário and Knorr-Held, 2003), spatial and spatio-temporal GMRF models (Gössl *et al.*, 2001), dynamic (generalized) linear models (Fahrmeir, 1992) and structured additive regression (Fahrmeir and Lang, 2001).

A particularly interesting feature of INLA is that it provides leave-one-out cross-validatory model checks without re-running the model for each observation in turn. In this paper we review the computation of the conditional predictive ordinate (CPO)

and the probability integral transform (PIT) in INLA and compare it with computation of the corresponding quantities using MCMC. We also consider posterior predictive model checks based on the whole data as an alternative to cross-validation. Section 2 reviews INLA and gives a detailed description how cross-validatory model checks are computed with INLA. Section 3 describes how these quantities are computed with MCMC. An extensive case study using an example from spatial epidemiology is described in Section 4 to compare the two approaches. We close with some discussion in Section 5.

2 The INLA Approach

The following section reviews INLA as an approach for approximate Bayesian inference in latent Gaussian models and shows how posterior and cross-validatory predictive checks can be computed using INLA.

2.1 Parameter Estimation with INLA

Consider a three-stage Bayesian hierarchical model based on an observation model $\pi(y|x) = \prod_i \pi(y_i|x_i)$, a parameter model $\pi(x|\theta)$, and a hyperprior $\pi(\theta)$. Here $y = (y_1, \ldots, y_n)$ denotes the observed data, x are unknown parameters which typically follow a GMRF, and θ are unknown hyperparameters. Note that reparametrization and parameter augmentation can be used to achieve $\pi(y_i|x) = \pi(y_i|x_i)$. The dimension of x will often be larger than n and we assume in the following that only the first n components of x are directly linked to the observations y.

Consider now the marginal posterior density

$$\pi(x_i|y) \quad = \quad \int_\theta \pi(x_i|\theta, y)\ \pi(\theta|y)\ d\theta$$

of the i-th component x_i of x. INLA approximates this by

$$\widetilde{\pi}(x_i|y) = \sum_k \widetilde{\pi}(x_i|\theta_k, y)\ \widetilde{\pi}(\theta_k|y)\ \Delta_k$$

using an approximation $\widetilde{\pi}(x_i|\theta, y)$ of $\pi(x_i|\theta, y)$ and an additional approximation $\widetilde{\pi}(\theta|y)$ of the marginal posterior density $\pi(\theta|y)$ of the hyperparameters θ. The weights Δ_k are chosen appropriately.

We first describe how $\pi(\theta|y)$ is approximated. Clearly,

$$\pi(x, \theta, y) = \pi(x|\theta, y) \times \pi(\theta|y) \times \pi(y), \tag{1}$$

so it follows that

$$\pi(\theta|y) \propto \frac{\pi(x, \theta, y)}{\pi(x|\theta, y)} \text{ for all } x.$$

INLA approximates $\pi(\theta|y)$ using a Laplace approximation (Tierney and Kadane, 1986):

$$\widetilde{\pi}(\theta|y) \propto \frac{\pi(x, \theta, y)}{\widetilde{\pi}_G(x|\theta, y)}|_{x=x^*(\theta)}.$$

The numerator can be easily evaluated based on (1). The denominator $\widetilde{\pi}_G(x|\theta, y)$ is the Gaussian approximation (Rue *et al.*, 2009, Section 2.2) of $\pi(x|\theta, y)$ and $x^*(\theta)$ is the mode of the full conditional $\pi(x|\theta, y)$, obtained through a suitable iterative algorithm.

The approximate posterior density $\widetilde{\pi}(\theta|y)$ is "numerically explored" to obtain suitable support points θ_k and the respective weights Δ_k.

For approximating the first component $\pi(x_i|\theta, y)$, a Gaussian approximation (Rue and Martino, 2007), easily extractable from $\widetilde{\pi}_G(x|\theta, y)$,

$$\tilde{\pi}_G(x_i|\theta, y) = \mathrm{N}(x_i; \mu_i(\theta), \sigma_i^2(\theta))$$

can be used. The approximation can be improved using a Laplace approximation

$$\tilde{\pi}_{LA}(x_i|\theta, y) \propto \mathrm{N}(x_i; \mu_i(\theta), \sigma_i^2(\theta)) \times \exp(\text{cubic spline}(x_i)),$$

or a simplified Laplace approximation based on the skew-normal distribution (Azzalini and Capitano, 1999), for details see Rue *et al.* (2009).

As suggested in Fahrmeir and Kneib (2009), it is instructive to compare the INLA approach with a REML/Empirical Bayes estimation in mixed models. In the empirical Bayes approach no hyperprior $\pi(\theta)$ is necessary, so the (RE)ML marginal likelihood corresponds to the marginal posterior $\pi(\theta|y)$. The (RE)ML marginal likelihood is maximized and only the (RE)ML estimate of θ is used, so no uncertainty with respect to θ is taken into account. The empirical Bayes estimate of x_i corresponds to the Gaussian approximation of $\pi(x_i|\theta, y)$ with θ fixed at the (RE)ML estimate. Hierarchical likelihood (Lee *et al.*, 2006) is a variation of this.

2.2 Posterior Predictive Model Checks with INLA

In order to check the fit of a Bayesian model posterior predictive checks were proposed by Gelman *et al.* (1996). The underlying concept of such checks is the posterior predictive distribution of a replicate observation Y_i which has density

$$\pi(y_i|y) = \int \pi(y_i|x_i, y) \cdot \pi(x_i|y)\, dx_i. \tag{2}$$

In Stern and Cressie (2000) it is suggested to use the posterior predictive p-value

$$\mathrm{Prob}(Y_i \leq y_i^{obs}|y)$$

as a measure of model fit, here y_i^{obs} denotes the actually observed count. If data are discrete, the posterior predictive mid-p-value (Berry and Armitage, 1995; Marshall and Spiegelhalter, 2003)

$$\mathrm{Prob}(Y_i < y_i^{obs}|y) + \frac{1}{2}\mathrm{Prob}(Y_i = y_i^{obs}|y)$$

can be used instead. An alternative quantity that may be of interest is the posterior predictive ordinate $\pi(y_i^{obs}|y)$. Small values of $\pi(y_i^{obs}|y)$ will indicate an outlying observation.

Extreme posterior predictive (mid-)p-values can be used to identify observations that diverge from the assumed model. However, one drawback concerning the interpretation of posterior predictive p-values is that they do not have a uniform distribution even if the data come from the assumed model. See Hjort *et al.* (2006), Marshall and Spiegelhalter (2007) and references therein for further details.

We will now explain how posterior p-values can be computed with INLA (Rue *et al.*, 2009). INLA returns an estimate of the posterior marginal of x_i in a discretised way: For $j = 1, ..., J$ support points $x_i^{(j)}$ an estimate $\widetilde{\pi}(x_i^{(j)}|y)$ of the posterior density

$\pi(x_i^{(j)}|y)$ is given. The support points are chosen such that they cover all areas with non-negligible posterior density. The value of the posterior predictive density (2) can then be approximated using the trapezoidal rule:

$$\hat{\pi}(y_i|y) \approx \sum_{j=2}^{J} \pi(y_i|\frac{1}{2}(x_i^{(j-1)} + x_i^{(j)})) \cdot \frac{1}{2}(x_i^{(j)} - x_i^{(j-1)})(\widetilde{\pi}(x_i^{(j)}|y) + \widetilde{\pi}(x_i^{(j-1)}|y)). \quad (3)$$

Of course, alternative techniques such as Simpson's rule can also be used. For discrete data, the posterior predictive (mid-)p-value can easily be derived as the sum of such probabilities. For $y_i = y_i^{obs}$ we obtain an estimate of the posterior predictive ordinate.

2.3 Leave-one-out Cross-validation with INLA

INLA routinely computes the DIC (Spiegelhalter *et al.*, 2002), a commonly used Bayesian model choice criterion. However, DIC may underpenalize complex models with many random effects (Plummer, 2008; Riebler and Held, 2009). Alternatively, the conditional predictive ordinate (CPO) (Pettit, 1990; Geisser, 1993) and the cross-validated probability integral transform (PIT) (Dawid, 1984) are available in INLA:

$$\begin{aligned} \text{CPO}_i &= \pi(y_i^{obs}|y_{-i}), \\ \text{PIT}_i &= \text{Prob}(Y_i \leq y_i^{obs}|y_{-i}). \end{aligned}$$

Here y_{-i} denotes the observations y with the i-th component omitted. This facilitates the computation of the cross-validated log-score (Gneiting and Raftery, 2007) for model choice. Similarly, PIT histograms (Czado *et al.*, 2009) can be computed to assess calibration of out-of-sample predictions.

We will now describe how these quantities are computed in INLA without rerunning the model. Throughout we assume that $y_{-i} = y_{-i}^{obs}$. However, we keep the explicit notation y_i^{obs} for the i-th observation to avoid confusion with other possible realisations of the corresponding random variable Y_i. As before, the vector y will always contain the observed data including y_i^{obs}.

First note that

$$\text{CPO}_i = \int \pi(y_i^{obs}|y_{-i}, \theta)\pi(\theta|y_{-i})d\theta, \quad (4)$$

$$\text{PIT}_i = \int \text{Prob}(Y_i \leq y_i^{obs}|y_{-i}, \theta)\pi(\theta|y_{-i})d\theta. \quad (5)$$

The first term in the integral in (4) now equals

$$\pi(y_i^{obs}|y_{-i}, \theta) = 1 \Big/ \int \frac{\pi(x_i|y, \theta)}{\pi(y_i^{obs}|x_i, \theta)} dx_i. \quad (6)$$

To see this, first note that

$$\pi(x_i|y_{-i}, \theta) = \frac{\pi(x_i|y, \theta)\pi(y_i^{obs}|y_{-i}, \theta)}{\pi(y_i^{obs}|x_i, \theta)}. \quad (7)$$

Integration with respect to x_i gives (6).

In practice, (6) is computed using numerical integration. The denominator of the ratio in the integral in (6) is the likelihood contribution of the i-th observation and known. However, only an approximation $\widetilde{\pi}(x_i|y, \theta)$ of the numerator $\pi(x_i|y, \theta)$ is

known using INLA, as described in Section 2.1. It depends on the accuracy of this approximation how accurate the numerical integration is. In particular, it may happen that the ratio $\widetilde{\pi}(x_i|y,\theta)/\pi(y_i^{obs}|x_i,\theta)$ is multimodal or tends to infinity for extreme values of x_i. It may also be difficult to locate the region of interest, i.e. the region with non-negligible contributions of $\pi(x_i|y,\theta)/\pi(y_i^{obs}|x_i,\theta)$. Such features are an artefact and a consequence of an imprecise approximation of the numerator $\pi(x_i|y,\theta)$ in the tails. Fortunately, INLA flags such problematic cases, for details see Section 4.

The first term in the integral in (5) can be written as

$$\text{Prob}(Y_i \leq y_i^{obs}|y_{-i},\theta) = \int \text{Prob}(Y_i \leq y_i^{obs}|x_i,\theta)\pi(x_i|y_{-i},\theta)dx_i.$$

The first term in this integral can be computed easily from the likelihood. The second term is available from (7) using $\pi(y_i^{obs}|y_{-i},\theta)$ as computed in (6). As before, $\pi(x_i|y,\theta)$ is available approximately through INLA.

Finally, we need to compute

$$\pi(\theta|y_{-i}) = \frac{\pi(\theta|y)\pi(y_i^{obs}|y_{-i})}{\pi(y_i^{obs}|y_{-i},\theta)}. \tag{8}$$

The denominator $\pi(y_i^{obs}|y_{-i},\theta)$ is known from (6). An approximation to $\pi(\theta|y)$ is available from Section 2.1. Therefore, the normalizing constant

$$\pi(y_i^{obs}|y_{-i}) = 1\Big/\int \frac{\pi(\theta|y)}{\pi(y_i^{obs}|y_{-i},\theta)}d\theta \tag{9}$$

of (8) can be approximately calculated as

$$\widetilde{\pi}(y_i^{obs}|y_{-i}) = 1\Big/\sum_k \frac{\widetilde{\pi}(\theta_k|y)}{\widetilde{\pi}(y_i^{obs}|y_{-i},\theta_k)}\Delta_k. \tag{10}$$

Here the θ_k's are support points of the approximate marginal posterior density $\widetilde{\pi}(\theta|y)$, which has been obtained in the first step of the INLA fitting procedure as described in Section 2.1. So the estimate $\widetilde{\pi}(y_i^{obs}|y_{-i})$ is the *weighted harmonic mean* of the $\widetilde{\pi}(y_i^{obs}|y_{-i},\theta_k)$'s, $k = 1,\ldots,K$, with weights $w_k = \widetilde{\pi}(\theta_k|y)\Delta_k$.

All terms appearing in (4) and (5) are now computed. Final approximation of PIT_i using (5) is based on support points θ_k as in (10) by replacement of the integral with a finite sum. Concerning CPO_i, note that (4) has been approximated already in (10), so the additional integration is not necessary.

3 Predictive Model Checks with MCMC

MCMC delivers samples $x^{(1)},\ldots,x^{(S)}$ from the posterior distribution $\pi(x|y)$. Similarly, samples $\theta^{(1)},\ldots,\theta^{(S)}$ from the posterior distribution $\pi(\theta|y)$ of the hyperparameters can be obtained on a routine basis. These samples are typically dependent, but suitable "thinning" can be be applied to obtain approximately independent samples.

3.1 Posterior Predictive Model Checks with MCMC

Within MCMC the posterior predictive p-values can be derived by drawing a replicate observation $Y_i^{(s)}$ for each of the $s = 1,...,S$ samples $x_i^{(s)}$ of the MCMC run and counting,

how many replicated observations are less than or equal to the actually observed count y_i^{obs}. For discrete data, the posterior predictive mid-p-value and the posterior predictive ordinate can be computed analogously.

If the likelihood $\pi(y_i|x_i)$ is available in closed form, an alternative approach is to average the likelihood across all samples $x_i^{(s)}$ from $\pi(x_i|y)$:

$$\widehat{\pi}(y_i|y) = \frac{1}{S}\sum_{s=1}^{S}\pi(y_i|x_i^{(s)}).$$

This technique is known as Rao-Blackwellization (Gelfand and Smith, 1990; Robert and Casella, 2004; Casella and Robert, 1996) and is typically more accurate than the approach based on replicates $Y_i^{(s)}$ from the predictive density. However, the Monte-Carlo error of the sample-based version is easier to assess so we have used this estimate in Section 4.

3.2 Leave-one-out Cross-validation with MCMC

Omitting the dependence on θ in (6) we obtain

$$\pi(y_i^{obs}|y_{-i}) = 1\Big/\int\frac{\pi(x_i|y)}{\pi(y_i^{obs}|x_i)}dx_i. \tag{11}$$

The immediate Monte-Carlo estimate of (11) is simply the harmonic mean of the likelihood values $\pi(y_i^{obs}|x_i)$,

$$\hat{\pi}(y_i^{obs}|y_{-i}) = 1\Big/\frac{1}{S}\sum_{s=1}^{S}\frac{1}{\pi(y_i^{obs}|x_i^{(s)})}, \tag{12}$$

evaluated at samples $x_i^{(1)},\ldots,x_i^{(S)}$ from $\pi(x_i|y)$. This estimate goes back at least to Gelfand (1996) and is very easy to use in MCMC applications. However, the harmonic mean can be numerically unstable and may not even follow a central-limit theorem (Newton and Raftery, 1994). This manifests itself by the occasional occurrence of a value $x_i^{(s)}$ with small likelihood $\pi(y_i^{obs}|x_i^{(s)})$ and hence large effect on the estimate (12). Indeed, Raftery (1996) has noted that the reciprocal of (12) may not even have finite variance.

However, for the computation of (mid-)p-values the value of $\pi(y_i|y_{-i})$ needs to be known for all $y_i \leq y_i^{obs}$. An importance sampling approach (Robert and Casella, 2004) can be adopted to compute $\pi(y_i|y_{-i})$ for any y_i, not necessarily equal to y_i^{obs}. First rewrite $\pi(y_i|y_{-i})$ as

$$\begin{aligned}\pi(y_i|y_{-i}) &= \int\pi(y_i|x_i)\pi(x_i|y_{-i})dx_i \\ &= \int\pi(y_i|x_i)\frac{\pi(x_i|y_{-i})}{\pi(x_i|y)}\pi(x_i|y)dx_i.\end{aligned}$$

The importance sampling estimate of $\pi(y_i|y_{-i})$ based on samples $x_i^{(1)},\ldots,x_i^{(S)}$ from $\pi(x_i|y)$ is hence

$$\hat{\pi}(y_i|y_{-i}) = \frac{\sum_{s=1}^{S}\pi(y_i|x_i^{(s)})w_i^{(s)}}{\sum_{s=1}^{S}w_i^{(s)}} \tag{13}$$

with importance weights

$$w_i^{(s)} = \frac{\pi(x_i^{(s)}|y_{-i})}{\pi(x_i^{(s)}|y)} \propto \frac{1}{\pi(y_i^{obs}|x_i^{(s)})},$$

compare Robert and Casella (2004, Equation (3.10)). For count data, the computation of cross-validatory (mid-)p-values reduces then to summing up the estimates $\hat{\pi}(y_i|y_{-i})$ for $y_i = 0, \ldots, y_i^{obs}$ (Marshall and Spiegelhalter, 2003). Note that the importance sampling estimate (13) reduces to the harmonic mean (12), if $y_i = y_i^{obs}$.

The variance of importance sampling estimators is difficult to assess; in fact the estimate may not even have finite variance. In particular, if the weights $w_i^{(s)}$ vary widely, they will give too much importance to only a few values of $\pi(y_i|x_i^{(s)})$ and the estimator (13) will be quite unstable, even for large S. However, we have investigated the weights $w_i^{(s)}$ in Section 4 and have found no weight particularly large relative to the others.

3.3 Approximate Cross-validation with MCMC

We now describe an alternative approach, based on an idea originally presented by Marshall and Spiegelhalter (2003) for approximate cross-validation in disease mapping models via MCMC. The method is based on the assumption that

$$\pi(\theta|y_{-i}) \approx \pi(\theta|y).$$

This assumption is plausible for moderate to large dimension of y, since θ is a *global* hyperparameter. Its posterior distribution based on all observations y should not change much if a single observation y_i is omitted.

The Marshall and Spiegelhalter (2003) *mixed predictive approach* is to generate additional samples

$$\tilde{x}_i^{(s)} \sim \pi(x_i|\theta^{(s)}, y_{-i})$$

$s = 1, \ldots, S$, where $\theta^{(s)}$ is a sample from $\pi(\theta|y)$. The samples $\tilde{x}_i^{(s)}$ do not directly depend on y_i, only indirectly because $\theta^{(s)} \sim \pi(\theta|y)$ does depend on y_i. The $\tilde{x}_i^{(s)}$'s are therefore approximately cross-validated and can be used in various ways to compute the predictive model checks discussed earlier.

A straightforward approach to compute PIT values is to draw additional samples $\tilde{y}_i^{(s)}$ from the pseudo-cross-validated predictive distribution and to compute the proportion of samples which are not larger than the observed value y_i^{obs}. Similarly, CPO values can be estimated based on the proportion of samples equal to y_i^{obs}. Alternatively a Rao-Blackwell approach as described in Section 3.1 can be used. In our application the PIT and CPO values resulting from the sampling strategy and the Rao-Blackwellization were almost identical. Mixed predictive PIT and CPO values shown in the following section are computed using Rao-Blackwellization.

4 Application

In our application we consider a typical example from spatial epidemiology. The data considered are cases of bovine viral diarrhoe (BVD) among cows in Switzerland collected during the year 2008. On behalf of an eradication program each cow in

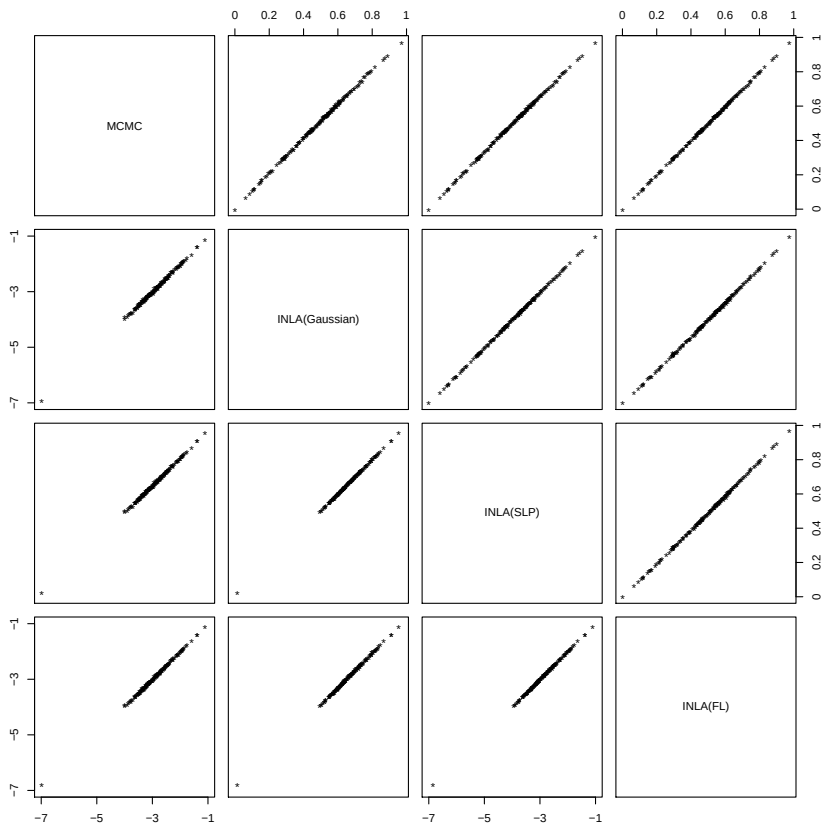


Figure 1: Scatterplots of posterior predictive mid-p-values (above diagonal) and log posterior predictive ordinates (below diagonal) computed by MCMC and INLA using the Gaussian (Gaussian), simplified Laplace (SLP) and full Laplace (FL) approximation

Switzerland was tested and the herd was marked as infected, if one or more diseased cows within this herd were detected. As Switzerland is divided in 184 administrative regions, the number of cases is available aggregated on regional level. Additionally, the Principality of Liechtenstein was included in the analysis. A number of 7164 cases was detected in total. For one region the number of cases is missing.

Under the rare disease assumption the usual starting point is to assume that the number of disease cases y_i in region $i = 1, \ldots, 185$ is Poisson distributed with parameter λ_i, which can be interpreted as the relative risk of the disease in the respective region. Additionally, the number of herds m_i is included in the model as an offset to adjust for the different number of herds living in each region. Using a standard formulation with Poisson observation model and a logarithmic link the relative risk parameter λ_i is modelled using the specification

$$\eta_i = \log(\lambda_i) = \log(m_i) + \psi_i + \nu_i. \tag{14}$$

The spatially unstructured component ν_i is assumed to be i.i.d. normally distributed with zero mean and unknown precision τ_ν whereas ψ_i is assumed to be structured in space. To account for the assumption that geographically close areas have simi-

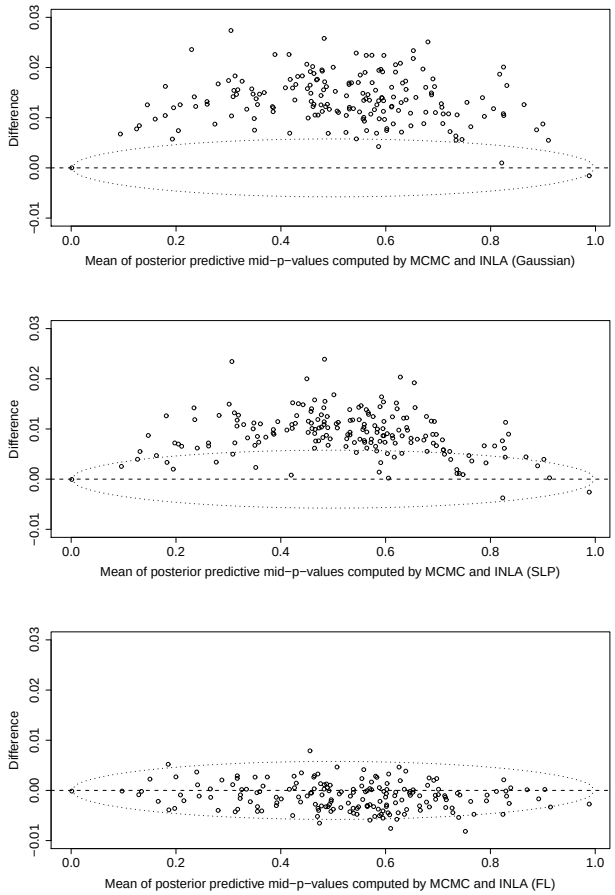


Figure 2: Bland-Altman plot to investigate the agreement between posterior predictive mid-p-values computed by MCMC vs. INLA using the Gaussian, simplified Laplace and full Laplace approximation. The dotted lines indicate pointwise 95%-confidence intervals based on the Monte-Carlo error attached to the MCMC estimates

lar incidence rates the spatially structured component ψ_i is modelled as an intrinsic Gaussian Markov random field with unknown precision τ_ψ (Rue and Held, 2005). This model was proposed by Besag *et al.* (1991), an extension to include covariates has been considered in Clayton and Bernardinelli (1992). The hyperpriors are chosen as $\tau_\psi \sim \text{Ga}(1, 0.018)$ and $\tau_\nu \sim \text{Ga}(1, 0.01)$, compare Bernardinelli *et al.* (1995) and Schrödle and Held (2010) for some motivation.

For the following analyses an MCMC run of length 930 000 was performed. Using every 30th iteration and a burn-in of 30 000 iterations, 30 000 MCMC samples have been stored. We also tested all three approximation methods available within INLA, as they are known to be differently accurate (Rue and Martino, 2007; Rue *et al.*, 2009). All calculations were done using the `inla` program version number 1.526.

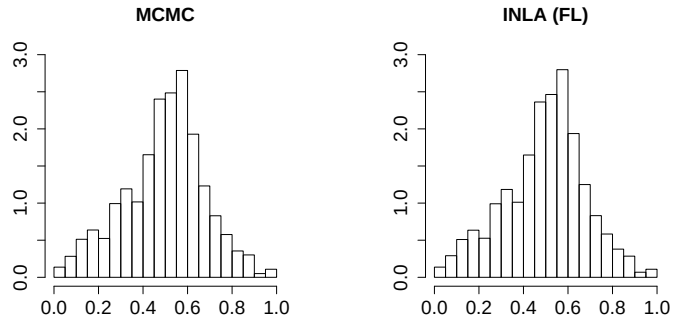


Figure 3: Adjusted histograms of posterior predictive p-values computed by MCMC and INLA using the full Laplace approximation

4.1 A Comparison of Posterior Predictive Model Checks

In the following the difference between the posterior predictive ordinates and posterior predictive mid-p-values computed by MCMC and INLA using three different approximation methods for the latent Gaussian field will be assessed.

Pairwise scatterplots are shown in Figure 1. The distribution of the posterior predictive ordinates is quite skewed and therefore shown on the log-scale. As can be seen from the plot, the estimates obtained with the four different methods look virtually identical.

The extent of agreement between any two methods can be visually examined in more detail using a plot suggested in Bland and Altman (1986), see also Kirkwood and Sterne (2003). The difference between two estimates is plotted on the vertical axis against the mean of each pair on the horizontal axis, see Figure 2. Also shown are 95%-confidence intervals indicating the Monte Carlo error attached to the MCMC estimates. The Monte Carlo standard error has been computed based on the assumption that the MCMC samples are independent. This assumption has been checked by visually inspecting the corresponding empirical autocorrelation functions.

Using this plot systematic bias can be detected and it can be examined if the differences between pairs of estimates depend on the actual value of the estimate. Posterior predictive mid-p-values obtained using the Gaussian and simplified Laplace approximation are slightly biased and typically smaller than the corresponding MCMC estimates. The bias is largest for mid-p-values around 0.5. For the full Laplace approximation the differences are close to zero and do not show any specific pattern. In fact, nearly all differences are now within the Monte Carlo confidence limits, i.e. the differences can be explained solely by the Monte Carlo error attached to the MCMC estimates. The MCMC estimates based on Rao-Blackwell were even closer to the INLA estimates.

Histograms of posterior predictive mid-p-values can be computed in analogy to the PIT histogram (Czado *et al.*, 2009), which was recently proposed for count data. The results are shown in Figure 3 based on MCMC and INLA using the full Laplace approximation. There is virtually no difference to see.

The histograms can be compared with histograms of the cross-validated PIT values in Figure 6. As mentioned in Stern and Cressie (2000) and Marshall and Spiegelhalter

Table 1: Number of unreliable CPO/PIT values for the Gaussian, simplified Laplace and full Laplace approximation

Gaussian	56 unreliable CPO/PIT values
Simplified Laplace	18 unreliable CPO/PIT values
Full Laplace	13 unreliable CPO/PIT values

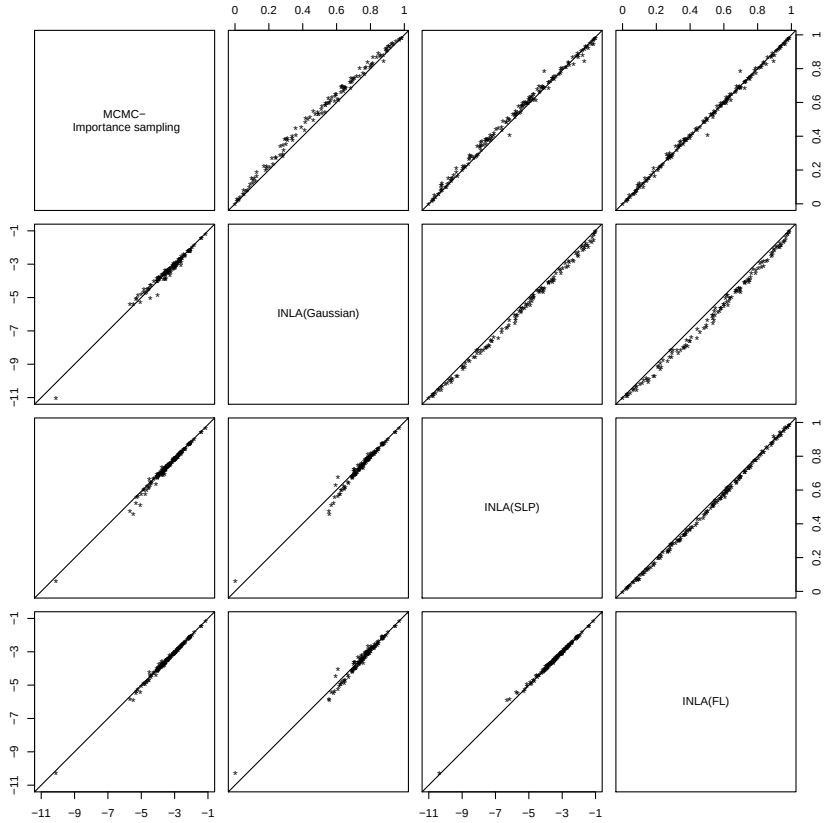


Figure 4: Scatterplots of leave-one-out cross-validatory predictive mid-p-values (above diagonal) and log conditional predictive ordinates (below diagonal) computed by MCMC vs. INLA using the Gaussian (Gaussian), simplified Laplace (SLP) and full Laplace (FL) approximation

(2007) posterior predictive p-values are not uniformly distributed and tend to be too conservative as the data are used twice. Indeed, the histograms in Figure 3 are far from uniformity with too many observations having mid-p-values around 0.5.

4.2 A Comparison of Leave-one-out Cross-validated Predictive Checks

Leave-one-out cross-validated predictive checks overcome the difficulties of posterior predictive checks mentioned in Section 4.1 and can be used to assess the predictive quality of a model (Marshall and Spiegelhalter, 2003; Czado *et al.*, 2009). Histograms

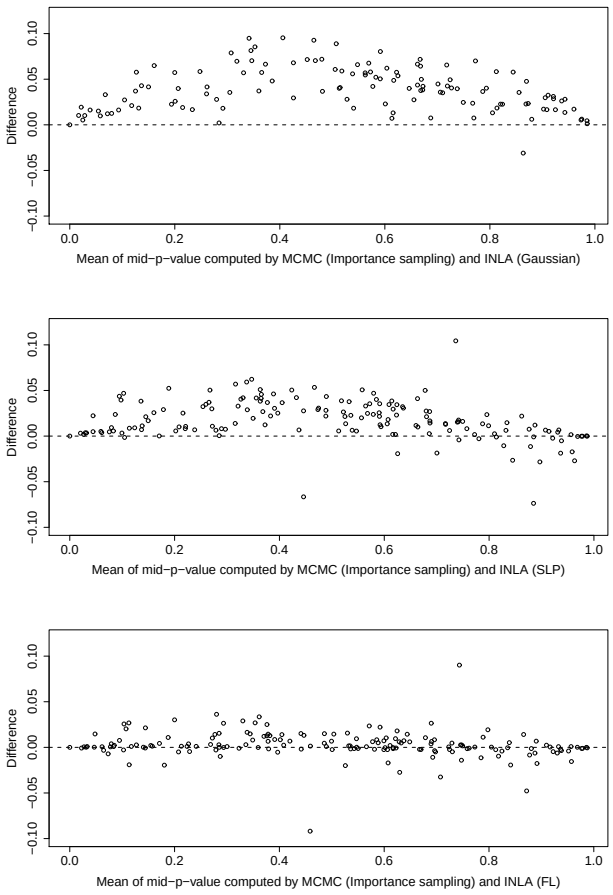


Figure 5: Bland-Altman plot to investigate the agreement between leave-one-out cross-validated mid-p-values computed by MCMC (importance sampling) vs. INLA using the Gaussian, simplified Laplace and full Laplace approximation

of the PIT values have been proposed to assess the calibration of a model (Czado *et al.*, 2009), the logarithmic score (Gneiting and Raftery, 2007), the sum of the log CPO values, can be used for model choice.

INLA returns the CPO and PIT values, as described in Section 2.3. Since the approximation methods for the latent Gaussian field are known to be differently accurate (Rue and Martino, 2007; Rue *et al.*, 2009), an empirical comparison is conducted. However, numerical problems may occur when CPO and PIT values are computed in INLA. Some of the CPO and PIT values might not be reliable due to numerical problems in evaluating the integral in (6). INLA automatically stores a file `failure.dat` which contains failure flags for each observation. We considered CPO/PIT values with flag equal to 1 as unreliable. Further details on this issue can be found in Martino and Rue (2009).

In Table 1 it is listed for how many observations the computation failed. Most failures occur based on the Gaussian approximation, the full Laplace approximation

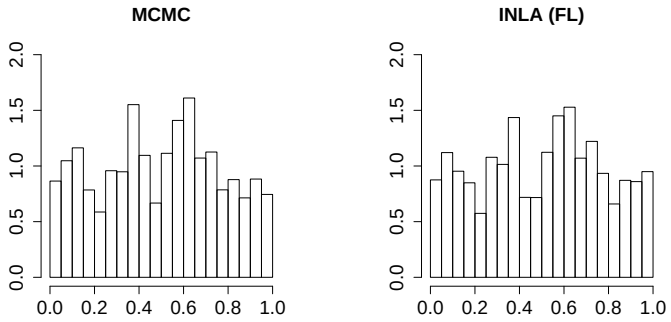


Figure 6: Adjusted histogram of PIT values computed by MCMC and INLA using the full Laplace approximation

performs best.

In order to assess the performance of INLA the output will be compared with results from a MCMC analysis based on the estimates (12) and (13). Mid-p- and log CPO values calculated with INLA and MCMC are shown in Figure 4. Each sub-figure is based on all those observations where CPO and PIT values could be computed without failure with the corresponding INLA approximation technique(s) considered.

Figure 4 reveals that the full Laplace approximation is closest to MCMC concerning bias and the differences between the full Laplace and the MCMC output do not show any specific pattern. More details can be seen on the corresponding Bland-Altman plots of the leave-one-out cross-validated mid-p-values, see Figure 5. First of all, a comparison with the corresponding plot showing the posterior predictive mid-p-values (Figure 2) reveals that the differences between MCMC and INLA have increased. However, a similar pattern as in Figure 2 can be seen, with mainly positive differences for the Gaussian and simplified Laplace approximation. In contrast, the mid-p-values computed with the full Laplace approach are closest to the MCMC estimates and do not exhibit a systematic bias. The corresponding PIT histograms are shown in Figure 6 and are quite similar. Note that the PIT histograms are much closer to a uniform distribution than the corresponding posterior predictive histograms shown in Figure 3.

4.3 A Comparison of Approximate Cross-validation with Posterior and Leave-one-out Predictive Checks using MCMC

CPO and mid-p-values resulting from a MCMC analysis have also been computed using the mixed predictive approach by Marshall and Spiegelhalter (2003) as described in Section 3.3. The approach is based on posterior samples of the precisions $\tau_v^{(s)}$ and $\tau_\psi^{(s)}$ based on the full data.

Approximately cross-validated samples of η_i and ψ_i are generated in a two-stage procedure based on a reparametrization of model (14) described in Knorr-Held and Rue (2002): First, $\tilde{\psi}_i^{(s)}$ is drawn from the conditional density

$$\tilde{\psi}_i^{(s)} | \psi_{-i}^{(s)}, \tau_\psi^{(s)} \sim N(\frac{1}{n_i} \sum_{j:j\sim i} \psi_j^{(s)}, \frac{1}{n_i \cdot \tau_\psi^{(s)}}).$$

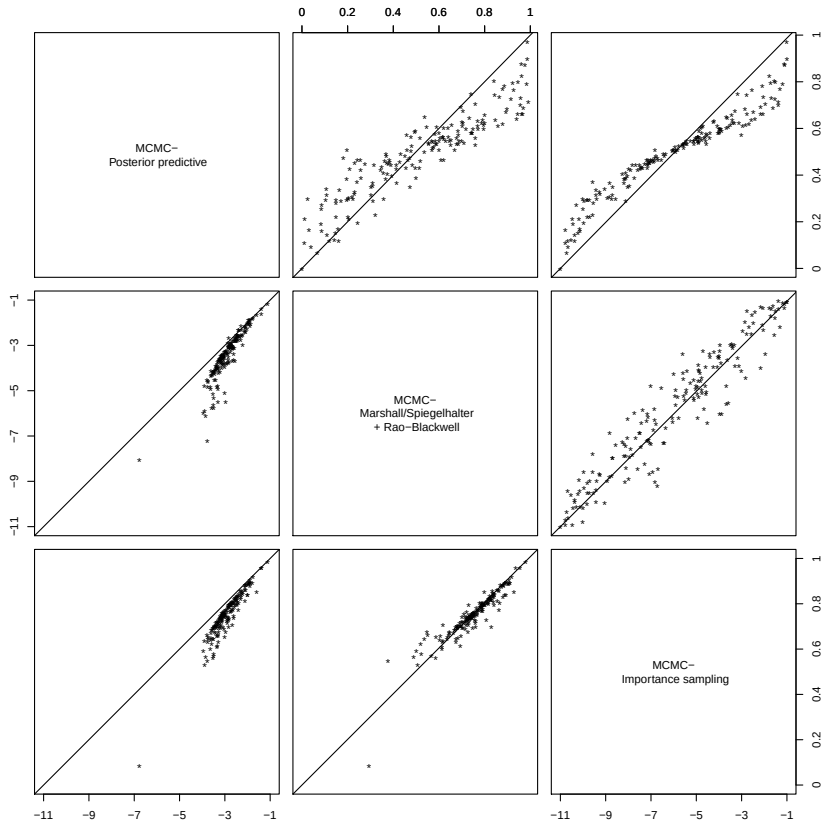


Figure 7: Scatterplots of mid-p- (above diagonal) and log CPO-values (below diagonal) computed by MCMC using three different approaches: The posterior predictive approach, the mixed predictive approach proposed by Marshall and Spiegelhalter in combination with a Rao-Blackwellization, and importance sampling

Here n_i denotes the number of neighbours of region i. In a second step, a sample $\tilde{\eta}_i^{(s)}$ of the linear predictor is drawn using

$$\tilde{\eta}_i^{(s)} | \tilde{\psi}_i^{(s)}, \tau_v^{(s)} \sim N(\tilde{\psi}_i^{(s)}, \frac{1}{\tau_v^{(s)}}).$$

This gives pseudo-cross-validated samples $\tilde{\eta}_i^{(s)}$ of the linear predictor, as proposed in Marshall and Spiegelhalter (2003).

Figure 7 compares the mixed predictive approach with the posterior predictive and the cross-validatory approach based on importance sampling. Compared with the importance sampling and the mixed predictive estimates, the posterior predictive estimates are systematically biased. As suspected, the mid-p-values are shrunk towards 0.5. Interestingly, the mixed predictive approach is closer to the ("exact") cross-validatory approach based on importance sampling. There is no systematic bias, although there is some variation in the estimates. This is in contrast to Marshall and Spiegelhalter (2003), who report that the mixed predictive approach performs better

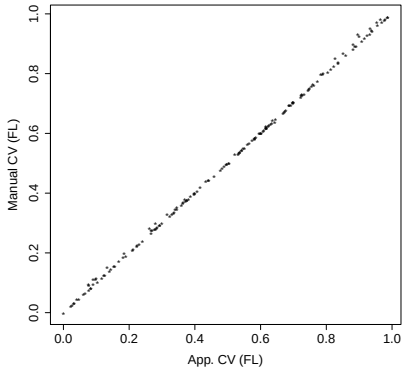


Figure 8: Scatterplot of manually computed mid-p-values using INLA vs. approximate mid-p-values obtained from the standard INLA output; the comparison was conducted for the full Laplace approximation

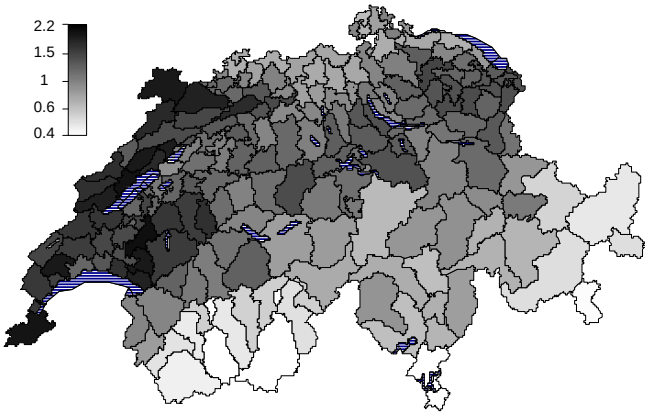


Figure 9: Fitted relative incidence of BVD in Switzerland, 2008

than the importance sampling approach in a similar disease mapping model using the well-known Scotland lip cancer data.

5 Discussion

The case study revealed that the cross-validatory checks provided by INLA are close to "exact" importance sampling estimates based on MCMC. The agreement is best if the full Laplace approximation is used. However, the relatively large number of failures is a drawback. Fortunately, these failures are flagged by INLA and it is straightforward to "manually" remove such an observation and to compute the desired leave-one-out quantities directly. The predictive distribution for the observation removed can be

calculated in exactly the same way as the posterior predictive distribution, see Section 2.2. For illustration, Figure 8 compares manually computed mid-p-values with the mid-p-values calculated based on the techniques described in Section 2.3 using the full Laplace approximation. The amount of agreement is remarkable.

We finally illustrate how the cross-validated log-score can be used for model comparison. To do so, we have considered two alternative models with either the unstructured or the structured component removed. The logarithmic score in the full model is -3.459, while in the reduced model with no unstructured component the score is even slightly larger (-3.454). However, the score of the model with only an unstructured component is considerably smaller (-3.779). This indicates that the structured component in the model is important, whereas the unstructured component can be omitted. The estimated relative incidence obtained from the best model without unstructured component is finally shown in Figure 9.

Acknowledgements

Financial support by the Swiss Federal Veterinary Office (BVET) is gratefully acknowledged.

References

Azzalini, A. and Capitano, A. (1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society: Series B* **61**: 579–602.

Bernardinelli, L., Clayton, D. and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors?, *Statistics in Medicine* **14**: 2411–2431.

Berry, G. and Armitage, P. (1995). Mid-p confidence intervals: a brief review, *Statistician* **44**: 417–423.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics* **43**(1): 1–59.

Bland, J. and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet* **i**: 307–310.

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**: 9–25.

Casella, G. and Robert, C. (1996). Rao-Blackwellisation of sampling schemes, *Biometrika* **83**: 81–94.

Clayton, D. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk, *in* J. Cuzick et al. (eds), *Geographical and Environmental Epidemiology. Methods for Small Area Studies*, Oxford University Press, pp. 205–220.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data, *Biometrics* **65**(4): 1254–1261.

Dawid, A. P. (1984). Statistical theory: The prequential approach, *Journal of the Royal Statistical Society: Series A* **147**: 278–292.

Fahrmeir, L. (1992). Posterior mode estimation by extended kalman filtering for multivariate dynamic generalized linear models, *Journal of the American Statistical Association* **87**: 501–509.

Fahrmeir, L. and Kneib, T. (2009). Discussion of Rue *et al.* (2009), *Journal of the Royal Statistical Society: Series B* **71**: 367.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society: Series C* **50**(2): 201–220.

Fahrmeir, L., Kneib, T. and Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective, *Statistica Sinica* **14**(3): 731–761.

Geisser, S. (1993). *Predictive Inference: An Introduction*, Chapman & Hall, London.

Gelfand, A. E. (1996). Model determination using sampling-based methods, *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov chain Monte Carlo in Practice*, Chapman & Hall, London, pp. 145–161.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**(410): 398–409.

Gelman, A., Meng, X.-L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies, *Statistica Sinica* **6**: 733–807.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**: 359–378.

Gössl, C., Auer, D. P. and Fahrmeir, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging, *Biometrics* **57**(2): 554–562.

Hjort, N. L., Dahl, F. A. and Steinbakk, G. H. (2006). Post-processing posterior predictive p-values, *Journal of the American Statistical Association* **101**(475): 1157–1174.

Kirkwood, B. and Sterne, J. (2003). *Medical Statistics*, 2nd edn, Blackwell Publishing, Oxford.

Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping, *Scandinavian Journal of Statistics* **29**(4): 597–614.

Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects - Unified Analysis via H-likelihood*, Chapman & Hall/CRC, London.

Marshall, E. C. and Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease-mapping methods, *Statistics in Medicine* **22**(4): 1649–1660.

Marshall, E. C. and Spiegelhalter, D. J. (2007). Identifying outliers in Bayesian hierarchical models: a simulation-based approach, *Bayesian Analysis* **2**(2): 409–444.

Martino, S. and Rue, H. (2009). Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the INLA program. `http://www.math.ntnu.no/~hrue/GMRFLib`.

Natário, I. and Knorr-Held, L. (2003). Non-parametric ecological regression and spatial variation, *Biometrical Journal* **45**(6): 670–688.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion), *Journal of the Royal Statistical Society: Series B* **56**: 3–48.

Pettit, L. I. (1990). The conditional predictive ordinate for the normal distribution, *Journal of the Royal Statistical Society: Series B* **52**: 175–184.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison, *Biostatistics* **9**(3): 523–539.

Raftery, A. E. (1996). Hypothesis testing and model selection, *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov chain Monte Carlo in Practice*, Chapman & Hall, London, pp. 163–187.

Riebler, A. and Held, L. (2009). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Technical report*. Conditionally accepted for *Biostatistics*.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer, Berlin.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC Press, London.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models, *Journal of Statistical Planning and Inference* **137**: 3177–3192.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society: Series B* **71**: 319–392.

Schrödle, B. and Held, L. (2010). Spatio-temporal disease mapping using INLA, *Environmetrics*. DOI: 10.1002/env.1065.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society: Series B* **64**: 583–639.

Stern, H. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models, *Statistics in Medicine* **19**: 2377–2397.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**(393): 82–86.

PAPER II

A primer on disease mapping and ecological regression using INLA

Birgit Schrödle & Leonhard Held

Pre-print version of a paper published in *Computational Statistics*, 2011, **26** (6), 241-258.

The original publication is available at

`http://www.springerlink.com/content/n158p80372434r7h/fulltext.pdf`.

A primer on disease mapping and ecological regression using INLA

Birgit Schrödle and Leonhard Held

Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Switzerland

Spatial and spatio-temporal disease mapping models are widely used for the analysis of registry data and usually formulated in an hierarchical Bayesian framework. Explanatory variables can be included by a so-called ecological regression. It is possible to assume both a linear and a non-parametric association between the disease incidence and the explanatory variable. Integrated nested Laplace approximations (INLA) can be used as a tool for Bayesian inference. INLA is a promising alternative to Markov chain Monte Carlo (MCMC) methods which provides very accurate results within short computational time. It is shown in this paper, how parameter estimates for well-known spatial and spatio-temporal models can be obtained by running INLA directly in R using the package INLA. Selected R code is shown. An emphasis is given to the inclusion of an explanatory variable. Cases of Coxiellosis among Swiss cows from 2005 to 2008 are used for illustration. The number of stillborn calves is included as time-varying covariate. Additionally, various aspects of INLA such as model choice criteria, computer time, accuracy of the results and usability of the R package are discussed.

Keywords: Disease mapping; Ecological regression; INLA; Spatio-temporal models

1 Introduction

Spatial and spatio-temporal disease mapping are widespread tools for passive surveillance of a disease and, therefore, used by epidemiologists on a standard basis. The most popular approach to spatial disease mapping was suggested in Besag *et al.* (1991) and developed further by several authors (Clayton and Bernardinelli, 1992; Bernardinelli *et al.*, 1995a). The methodology can be extended to the spatio-temporal case by inclusion of a linear (Bernardinelli *et al.*, 1995b; Assunção *et al.*, 2001) or nonparametric trend in time and time-space interactions (Knorr-Held, 2000; Schmid and Held, 2004). In order to investigate the association of an explanatory variable with the geographical and temporal variation in disease risk, a so-called ecological regression model can be built (Clayton *et al.*, 1993). The effect of the covariate can be modelled in a linear or a nonparametric fashion (Fahrmeir and Lang, 2001; Natario and Knorr-Held, 2003).
Such spatial and spatio-temporal disease mapping models are usually formulated in a hierarchical Bayesian framework with a latent Gaussian Markov random field (GMRF) (Clayton and Bernardinelli, 1992; Rue and Held, 2005). So far, Markov chain Monte Carlo (MCMC) techniques have been used for Bayesian inference, but these techniques

are very time-consuming since spatio-temporal disease mapping models form a complex class. Elaborate MCMC algorithms have to be used to obtain reliable posterior estimates and the MCMC output might be hard to interpret for the standard user. Integrated nested Laplace approximations (INLA) have recently been proposed as a promising alternative (Rue *et al.*, 2009). The methodology offers very accurate approximations of the posterior marginals in short computational time. Additionally, a tool for Bayesian model choice, namely the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002), and predictive measures as the logarithmic score (Gneiting and Raftery, 2007) and the probability integral transform (PIT) (Czado *et al.*, 2009) can be obtained.
The INLA approach is easy to apply, since a `C` program called `inla` is available. Furthermore, the `inla` program is bundled within an `R` package `INLA` to improve usability as a standard tool. As the INLA approach is a complex numerical procedure, it might still not be easy for the user to choose the right specifications and features. Hence, the `R` code needed for inference in spatial and spatio-temporal models and ecological regression using `INLA` will be introduced and the usability of the approach will be discussed.
The paper is organized as follows: First the INLA methodology is introduced briefly and possible options for the approximation algorithms are shown. In Section 3 some well-known spatial and spatio-temporal models for disease mapping are applied to reported Coxiellosis cases among Swiss cows from 2005 to 2008 using `INLA`. Additionally, an ecological regression analysis is performed including the number of stillborn calves as explanatory variable. In the subsequent section it is shown how the obtained output can be interpreted and used for model choice. Here the emphasis will be on spatio-temporal models. A look on computational issues and usability of the INLA approach will be taken in Section 4.3. A brief discussion is given in Section 5.

2 INLA

Spatial and spatio-temporal models as will be introduced in Section 3 are built as Bayesian hierarchical models with three stages: The first stage is the observational model $\pi(\mathbf{y}|\mathbf{x})$, where $\boldsymbol{y}$ denotes the observations. The vector $\boldsymbol{x}$ contains all components of the latent Gaussian field (GMRF) $\pi(\boldsymbol{x}|\boldsymbol{\theta})$. The GRMF is typically controlled by a few hyperparameters $\boldsymbol{\theta}$, which form the third stage. Their respective prior distribution is denoted as $\pi(\boldsymbol{\theta})$. The desired posterior marginals

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \pi(x_i|\boldsymbol{\theta},\boldsymbol{y})\ \pi(\boldsymbol{\theta}|\boldsymbol{y})\ d\boldsymbol{\theta}$$

of all components of the GMRF are approximated by INLA using the finite sum

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\boldsymbol{\theta}_k,\boldsymbol{y})\ \widetilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y})\ \Delta_k, \tag{1}$$

where $\widetilde{\pi}(x_i|\boldsymbol{\theta},\boldsymbol{y})$ and $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ denote approximations of $\pi(x_i|\boldsymbol{\theta},\boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, respectively. This finite sum is evaluated at support points $\boldsymbol{\theta}_k$ using appropriate weights Δ_k. The $\boldsymbol{\theta}_k$'s can be obtained in two different ways, see below.
From $\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \times \pi(\boldsymbol{\theta}|\boldsymbol{y}) \times \pi(\boldsymbol{y})$ it follows that the posterior marginal $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of the hyperparameters can be obtained using a Laplace approximation

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\Big|_{\boldsymbol{x}=x^*(\boldsymbol{\theta})}$$

(Tierney and Kadane, 1986), where the denominator $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ denotes the Gaussian approximation of $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ (Rue and Held, 2005). Gaussian approximation means that the distribution of a non-normal variable is approximated by a normal distribution by matching the mode and the curvature at the mode (Rue and Held, 2005, Section 4.4.1). According to Rue *et al.* (2009) it is sufficient to "numerically explore" this approximate posterior density using suitable support points $\boldsymbol{\theta}_k$ for (1). The first strategy is called GRID strategy and is computationally intensive. The mode of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ has to be found by some quasi-Newton method. Subsequently, the density around the mode is explored and points where the probability mass is considered as significant are selected for the integration. If the dimension h of hyperparameters included in the model is moderate (h=6-12), it is computationally more efficient to use the so-called central composite design (CCD) to lay out support points in the h-dimensional space. Here, centre points are augmented with a group of star points which allow for estimating the curvature of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$. For more details on both methods see Rue *et al.* (2009). As the CCD integration scheme needs much less computational time and the differences between CCD and GRID strategy are minor, Rue *et al.* (2009) recommend the use of the CCD strategy for problems with high dimensionality of the hyperparameter vector $\boldsymbol{\theta}$. The difference in computer time and the resulting marginals for both strategies are briefly discussed in Section 4.3.

To approximate the first component of (1), namely the posterior marginal for x_i conditioned on selected values of $\boldsymbol{\theta}_k$, three different approaches are possible: A Gaussian, a full Laplace and a simplified Laplace approximation. The Gaussian approximation is fastest, but according to Rue and Martino (2007) there can be errors in the location of the posterior marginals, errors due to the lack of skewness, or both. The Gaussian approximation can be improved by using a Laplace approximation (Tierney and Kadane, 1986), but this strategy is rather time-consuming. Hence, Rue *et al.* (2009) introduce the so-called simplified Laplace approximation which is less expensive from a computational point of view with only a slight loss of accuracy.

To run INLA, a `C` program called `inla` is offered by the authors of Rue *et al.* (2009), which performs all required computations in a modular way. This program is based on the `GRMFLib`-library, which incorporates efficient algorithms for sparse matrices (Rue and Held, 2005). Additionally, the computations are speeded up by the implementation of parallel computing elements. An `R`-interface called `INLA` is available to ease the usage of the `inla` program. The `inla` program is bundled within this `R` library (R Development Core Team, 2005). The software can be downloaded from `http://www.r-inla.org` and is running in a Linux, MAC and Windows environment. For the analyses within this paper we used the `INLA` library built on the 28th of April 2010. The respective `R` code is shown where it was considered as helpful. The data and further `R` code can be found within Online Resource 1.

3 Review of spatial and spatio-temporal disease mapping models and their specification using `INLA`

The following sections give an introduction to the data and show the specification of selected spatial and spatio-temporal disease mapping models using `INLA`.

Table 1: Number of reported cases of Coxiellosis in cows per year, 2005-2008.

Year	2005	2006	2007	2008
n	30	45	54	61

3.1 Data - Cases of Coxiellosis among cows in Switzerland, 2005-2008

Coxiellosis is a widespread infectious, endemic disease caused by the bacterium *coxiella burnetii* among ruminant animals (Aitken, 1989). In most cases it is subclinical, but it can be the reason for an abortion in a late phase of the pregnancy or a stillbirth (Woldehiwet, 2004). The spread of the bacterium can take place through ticks, but happens as well from animal to animal by airborne infection as the bacterium is present in abortion products and excreted by diseased animals in their milk, urine and excrement. Special attention must be payed to this disease as it is a so-called zoonosis that can also affect humans (Q fever); such epidemics have been observed in Switzerland (Dupuis *et al.*, 1987).
The data considered are cases of Coxiellosis among cows reported to the Swiss Federal Veterinary Office from 2005 to 2008. A herd is marked as infected, if one or more diseased cows were detected. The number of cases is available on a yearly basis for 184 regions of Switzerland. Additionally, data from the Principality of Liechtenstein is included. As shown in Table 1, the number of reported cases has constantly been rising during the last four years. Hence, it is of interest if a significant rise in reported cases took place and what the spatial distribution of the disease within Switzerland looks like.
As Coxiellosis is a widespread disease, it is obvious from Table 1 that massive underreporting must be present, although the disease is notifiable in Switzerland. Switzerland is a confederation of 26 cantons, which consist of one or more regions. The cantonal veterinary authorities are responsible for the realization of federal veterinary legislation in each affiliated region. In Schrödle *et al.* (2011) it was found that the number of reported cases within one region might depend on the canton it belongs to. Hence, cantons are considered as a second, coarser spatial grid.
In Section 3.2 the cases from 2008 only are used as response variable for spatial disease mapping, while the spatio-temporal disease mapping in Section 3.3 is illustrated using all cases from 2005 to 2008.

3.2 Spatial disease mapping

Under the rare disease assumption it is usually assumed that the number of disease cases y_i in region $i = 1, \ldots, 185$ is Poisson distributed with parameter λ_i, which can be interpreted as the relative risk of the disease in the respective region. Additionally, the number of herds m_i is included in the model as an offset to adjust for the different number of herds at risk. In the standard formulation established by Besag *et al.* (1991) the relative risk parameter λ_i is specified as

$$\eta_i = \log(\lambda_i) = \log(m_i) + \mu + \nu_i + \psi_i. \tag{2}$$

This model will be called BYM1. It contains a spatially unstructured component ν_i (variable name (vn): `region.nu`) which is i.i.d. normally distributed with zero mean and unknown precision τ_ν, whereas ψ_i (vn: `region.psi`) is assumed to be structured

in space. To account for the assumption that geographically close areas have similar incidence rates the spatially structured component ψ_i is modelled as an intrinsic Gaussian Markov random field (IGMRF) with unknown precision τ_ψ (Rue and Held, 2005). This specification is also called a conditionally autoregressive (CAR) prior (Banerjee *et al.*, 2004). To ensure identifiability of the intercept μ a sum-to-zero constraint must be imposed on the ψ_i's. The variables `region.nu` and `region.psi` are identical, but two different objects have to be specified in `INLA`.
As discussed in Schrödle *et al.* (2011), the number of reported cases per region might depend on the canton a region belongs to. For investigation of this fact, (2) is extended to a second, coarser spatial level: An i.i.d. random effect α_j (vn: `canton.alpha`) for each of the 26 cantons of Switzerland and Liechtenstein is added ($j = 1, ..., 27$). The resulting linear predictor is

$$\eta_i = \log(\lambda_i) = \log(m_i) + \mu + \nu_i + \psi_i + \alpha_{j(i)}. \tag{3}$$

The extended model will be called BYM2.
The choice of hyperpriors for disease mapping models is discussed in Bernardinelli *et al.* (1995a). As proposed we use Prior B, $\text{Ga}(1, 0.01)$ (vn: `prior.nu`), as hyperprior for τ_ν and τ_α. The prior for τ_ψ was adjusted for the structure of the Swiss graph and chosen as $\text{Ga}(1, 0.018)$ (vn: `prior.psi`) (Bernardinelli *et al.*, 1995a).
To run these models in `INLA`, the linear predictor of the model has to be specified as a `formula` object in `R` using the function `f()` for smooth effects. Subsequently, the specified model can be run using `inla()`.
The type of a smooth effect can be specified in `f()` using, e.g., `model="iid"` for an i.i.d. random effect and `"besag"` for an IGMRF like ψ. The respective graph-file (e.g. `"switzerland.graph"`) containing the neighbourhood structure has to be specified as well. The hyperpriors for the precision parameters of the smooth effects (argument: `param`) have to be chosen and linear constraints can be set (argument: `constraint`). For the `"besag"`-prior a sum-to-zero constraint is imposed as default. Within the `inla()`-call further options for the INLA algorithm can be set. Here it can, e.g., be specified if quantities for predictive measures (`cpo=1`) and the DIC (`dic=1`) should be computed and which strategy for the approximation of the latent Gaussian field and the posterior marginals of the hyperparameters $\boldsymbol{\theta}$ should be used. The default choice is the simplified Laplace approximation (SLP) and the CCD strategy. As shown in Held *et al.* (2010), the accuracy of the SLP approximation is often not sufficient for the computation of predictive measures. Hence, the full Laplace approximation was chosen in the following application (`strategy="laplace"`). A dataframe can be specified using the argument `data`; the offset (vn: `offset`) for a Poisson model is given to `INLA` via `E`. The vector `Y.cox` of the dataset `cox.08` contains the number of Coxiellosis cases per region in 2008. For more details see the `inla` manual (Martino and Rue, 2009).
The resulting model specification and the call to fit model BYM1 is

```
> f.BYM1<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
+              +f(region.psi,model="besag",param=prior.psi,
+                 graph.file="switzerland.graph")
> BYM1<-inla(f.BYM1,family="poisson",E=offset,data=cox.08,
+            control.inla=list(strategy="laplace"),
+            control.compute=list(dic=1,cpo=1))
```

Using `names(BYM1)` the components of the output can be seen. For example,

```
> round(BYM1$summary.random$region.nu[1, ], digits = 4)

  ID   mean     sd 0.025quant 0.5quant 0.975quant    kld
1  0 -0.283 1.0572    -2.5108  -0.2372     1.6483 0.0028
```

returns the results for the unstructured effect of region 1. Standard outputs are the posterior mean, standard deviation, 2.5%-, 50%- and 97.5%-quantiles and the symmetric Kullback-Leibler distance (SKLD) between the Gaussian and the (full) Laplace approximation, which is derived from the Kullback-Leibler discrepancy (KLD). The KLD is a measure to quantify the divergence between two density functions, but it is not symmetric (Kullback and Leibler, 1951). To solve this problem the SKLD is defined as the sum of the KLD's measured in both directions (Wood and Kohn, 1998; Moreno *et al.*, 2004). Model BYM2 can be specified in a similar fashion; the respective results are discussed in Section 4.1.

3.3 Spatio-temporal disease mapping

To find out, if there has been a statistically significant linear rise in reported cases of Coxiellosis from 2005 to 2008, a spatio-temporal disease mapping model is adopted in the following section. This model is analogous to Bernardinelli *et al.* (1995b), but expanded by a cantonal effect α_j as case reporting might be biased with regard to the cantonal affiliation of a region. The linear predictor can be written as

$$\eta_{it} = \log(m_i) + \mu + \nu_i + \psi_i + \alpha_{j(i)} + (\beta_1 + \delta_{j(i)}) \cdot t. \tag{4}$$

This model includes the same components as (3), but a main linear time trend β_1 (vn: `time.beta1`) and a so-called differential trend δ_j (vn: `differential.delta`) for each canton are added. The effect δ_j is modelled as a random slope and accounts for cantonal departures from the main linear time trend. As it is necessary to allow for correlation between intercept and slope in a random slope model, it is assumed that $(\alpha_j, \delta_j)^T$ follows a bivariate normal distribution with zero mean and some unknown precision matrix, to which a Wishart prior is assigned (Bernardinelli *et al.*, 1995b). Using `INLA` $(\alpha_j, \delta_j)^T$ can be defined using two components `model="2diidwishartpart0"` and `"2diidwishartpart1"`, respectively. Four parameters have to be specified within the `"2diidwishartpart0"`-component. In this application these parameters are chosen as `prior.wishart=c(4,1,1,0)`; this choice was used in Schrödle *et al.* (2011) for a similar setting and checked for sensitivity. For the differential trend δ_j appropriate weights (vn: `time`) given by the timepoints have to be introduced additionally.
This model is called ST1 and the model formula is defined as

```
> f.ST1<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
+              +f(region.psi,model="besag",param=prior.psi,
+                 graph.file="switzerland.graph")+
+              +f(canton.alpha,model="2diidwishartpart0",
+                 param=prior.wishart)+
+              +f(differential.delta,time,
+                 model="2diidwishartpart1")+
+              +time.beta1
> ST1<-inla(f.ST1,family="poisson",E=offset,data=cox,
+            control.inla=list(strategy="laplace"),
+            control.compute=list(dic=1,cpo=1))
```

Table 2: Number of stillborn calves per year.

Year	2005	2006	2007	2008
n	15326	23044	25289	26911

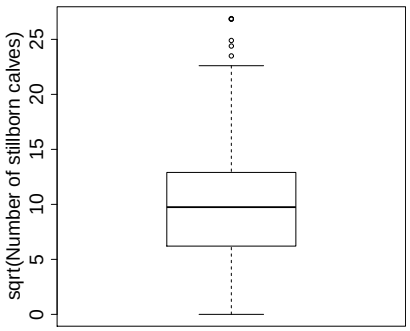


Figure 1: Boxplot of the square root of the number of stillborn calves, 2005-2008.

The vectors `Y.cox`, `offset`, `region.psi`, `region.nu` and `canton.alpha` in the dataset `cox` are four times as long as the corresponding vectors in the dataset `cox.08` from Section 3.2, since four years are included in the spatio-temporal analysis. The vectors `canton.alpha` and `differential.delta` are identical, but two different variables must be specified within `INLA`.
Another option would be to assume a nonparametric trend in time as proposed in Knorr-Held (2000). This approach was not considered here as only a short time interval is taken into account, but its implementation in `INLA` is possible (Schrödle *et al.*, 2011).

3.4 Ecological regression

The model in the preceding section can be extended to an explanatory variable to investigate its association with the geographical and temporal variation in disease risk (Clayton and Bernardinelli, 1992; Clayton *et al.*, 1993).
As noted in Section 3.2, case reporting in Switzerland might be biased by factors like, e.g., temporally varying disease awareness. So it is of interest, if the rise in reported cases (see Table 1) can be accounted to a "real" rise of disease incidence. Since Coxiellosis can cause the stillbirth of a calf (Aitken, 1989), a spatial and temporal association between the number of stillborn calves and reported Coxiellosis cases within one region would indicate a "real" rise in the incidence of the disease. The number of stillborn calves is available for each region and year and has constantly been growing since 2005, see Table 2. The covariate was square-root transformed before the analysis (vn: `ncalves.beta2`); a boxplot of the respective values can be found in Figure 1.

In Clayton and Bernardinelli (1992) and Clayton *et al.* (1993) it has been suggested to assume a linear relationship for the explanatory variable. Hence, (4) is expanded by

inclusion of a linear covariate z_{it}

$$\eta_{it} = \log(m_i) + \mu + \nu_i + \psi_i + \alpha_{j(i)} + (\beta_1 + \delta_{j(i)}) \cdot t + \beta_2 \cdot z_{it}. \quad (5)$$

This model will be denoted as model ST2.
Natario and Knorr-Held (2003) have proposed to replace the linear effect of z_{it} with a smooth nonparametric function f_z. The resulting model can be written as

$$\eta_{it} = \log(m_i) + \mu + \nu_i + \psi_i + \alpha_{j(i)} + (\beta_1 + \delta_{j(i)}) \cdot t + f_z(z_{it}). \quad (6)$$

This model will be called ST3.
In the easy case the covariate z_{it} can take only K equally spaced levels $g_1 < ... < g_k < ... < g_K$. Then $\gamma_k = f_z(z_{it} = g_k)$ is assumed to follow a random walk of second order on regular locations with joint density

$$\pi(\gamma|\tau_\gamma) \propto \exp\Big(-\frac{\tau_\gamma}{2}\sum_{k=2}^{K-1}(\gamma_{k+1} - 2\gamma_k + \gamma_{k-1})^2\Big) \quad (7)$$

and precision τ_γ (Fahrmeir and Lang, 2001; Rue and Held, 2005). This is a natural assumption, as a random walk of second order models deviations from a linear trend (Natario and Knorr-Held, 2003). In addition, it is appropriate for representing smooth curves with a small curvature and is computationally convenient due to its Markov property. A sum-to-zero constraint has to be imposed on the γ_k's to ensure identifiability of μ. As the levels of the covariate in this application are not equally spaced and the use of equal spaces would increase the dimension of the model (Lindgren and Rue, 2008), (7) has to be extended to the more general case with non-equally spaced levels. In Fahrmeir and Lang (2001) it has been suggested to include appropriate weights, but this approach leads to inconsistencies regarding variances for the case of non-equally spaced levels (Lindgren and Rue, 2008). Hence, a new approach has been proposed in Lindgren and Rue (2008), where (7) is interpreted as an approximated Galerkin solution to the stochastic differential equation $f''(t) = \mathrm{d}W(t)/\mathrm{d}t$, where $W(t)$ is a Wiener process (Rue and Held, 2005). This approach does not show inconsistencies regarding the variances and its covariance properties converge to those of a continuous RW2 process as the grid of the observed levels gets more dense. It is computationally convenient with negligible errors and, hence, implemented in `INLA` when using option `model="rw2"`. Therefore, model (6) can be specified using

```
> f.ST3<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
+             +f(region.psi,model="besag",param=prior.psi,
+                graph.file="switzerland.graph")+
+             +f(canton.alpha,model="2diidwishartpart0",
+                param=prior.wishart)+
+             +f(differential.delta,time,
+                model="2diidwishartpart1")+
+             +time.beta1+
+             +f(ncalves.gamma,model="rw2",param=prior.gamma)
```

Within `INLA` it is also possible to model γ as a continuous time random walk of second order (Rue and Held, 2005). This approach might be more time-consuming compared to the discretized approach in Lindgren and Rue (2008). To run it, the option `"rw2"` has to be replaced by `"crw2"`.
Care has to be taken concerning the prior chosen for the variance $\sigma_\gamma^2 = 1/\tau_\gamma$, as its

interpretation depends on the levels taken by the covariate and the distance ξ between successive values of g_k. As noted in Berzuini and Clayton (1994), the ratio of the prior mode of σ_γ, which specifies the prior belief in smoothness, and the squared distance ξ^2 should be kept constant when varying the parameters of the hyperprior. Usually an inverse gamma distribution IGa(a,b) with prior mode $b/(a+1)$ is adopted as hyperprior for σ_γ^2. In Natario and Knorr-Held (2003) it is recommended to use an IGa$(1, 0.00005)$ prior (vn: `prior.gamma`) for non-equally spaced covariates with an average distance of 1. Hence, the values of the covariate (vn: `ncalves.gamma`) in this application were scaled in a way that this requirement is satisfied. In Natario and Knorr-Held (2003) it was found that the nonparametric trend is sensitive to the choice of the prior; for the application at hand this will be investigated in more detail in Section 4.

4 Results and model choice

The following sections show, how the `INLA` output for all models specified in Section 3 can be used for model choice and interpretation. To shorten considerations, Section 4.1 deals only with model choice whereas in Section 4.2 results for the spatio-temporal analysis are presented as well. Some issues with regard to computer time and the use of different approximation techniques are briefly discussed in Section 4.3.

4.1 Spatial disease mapping - Model choice

As noted before, several quantities for model choice and model calibration are available by `INLA`. In order to decide which model provides the best trade-off between model fit and complexity, the DIC is given as a well-known Bayesian model choice criterion (Spiegelhalter *et al.*, 2002). The conditional predictive ordinates (CPO's) which facilitate the computation of the cross-validated logarithmic score for model choice (Gneiting and Raftery, 2007) are given as well as the probability integral transform (PIT), which can be used to assess calibration of out-of-sample predictions (Czado *et al.*, 2009). The use of these measures is exemplary shown for model choice between BYM1 and BYM2.
The DIC is the sum of a measure of model fit, the posterior `mean of the deviance` $\bar{D}$, and model complexity, the `effective number of parameters` p_D, and is addressed using

```
> BYM1$dic
```

```
                                  [,1]
mean of the deviance          158.37680
deviance of the mean          125.12526
effective number of parameters  33.25154
dic                           191.62834
```

```
> BYM2$dic
```

```
                                  [,1]
mean of the deviance          158.27891
deviance of the mean          130.09263
effective number of parameters  28.18627
dic                           186.46518
```

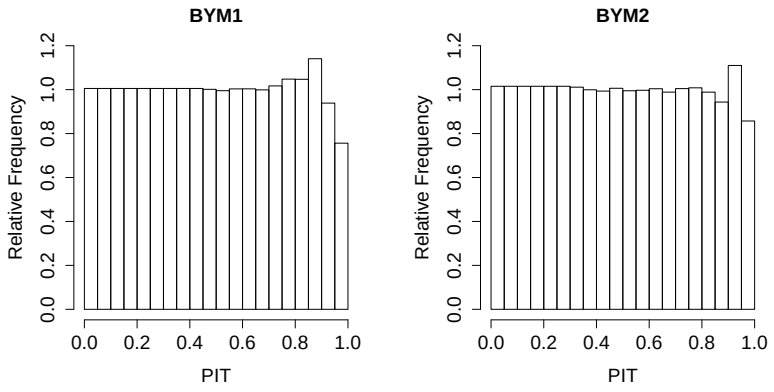


Figure 2: Adjusted PIT-histograms for models BYM 1 and BYM 2.

The smaller the DIC, the better the trade-off between model fit and complexity. The posterior deviance and the number of effective parameters in model BYM1 are slightly larger than in model BYM2. Hence, the DIC value of model BYM2 is smaller. The logarithmic score (Gneiting and Raftery, 2007) can be computed as

```
> lsBYM1 <- -mean(log(BYM1$cpo))
> lsBYM2 <- -mean(log(BYM2$cpo))
> round(lsBYM1, digits = 3)

[1] 0.55

> round(lsBYM2, digits = 3)

[1] 0.532
```

The smaller the resulting score, the better the predictive quality of the model. As the score for model BYM2 is a bit smaller the predictive quality for BYM2 is better. The calibration of both models can be checked by plotting an adjusted PIT histogram as suggested by Czado *et al.* (2009) using the values provided in, for example, `BYM1$pit`. The results can be seen in Figure 2; the histograms are close to uniform. Both models are almost perfectly calibrated, but the calibration of model BYM2 seems to be slightly better. Hence, BYM2 is preferred.

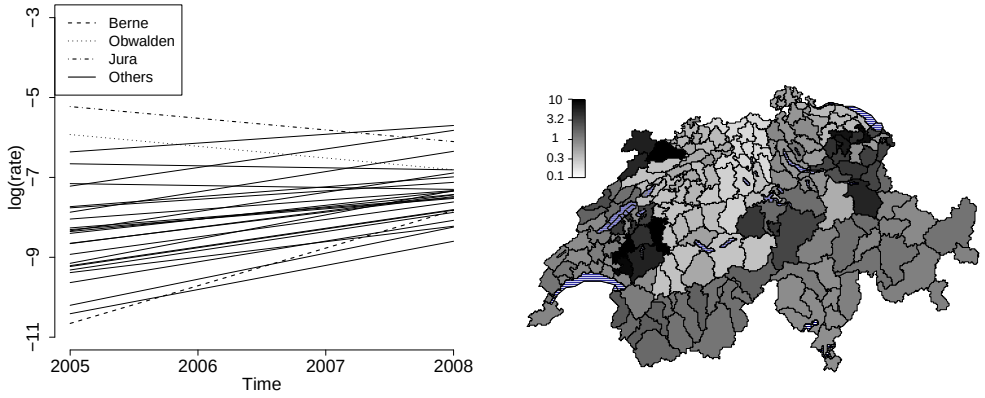
4.2 Spatio-temporal disease mapping - Results and model choice

The results for model ST1 are called using

```
> round(ST1$summary.fixed, digits = 4)

              mean     sd 0.025quant 0.5quant 0.975quant    kld
(Intercept) -7.6492 0.2780    -8.2219  -7.6395    -7.1272 0.5287
time.beta1   0.3400 0.1548     0.0514   0.3342     0.6637 0.0573

> round(ST1$summary.hyperpar[, c(1, 2)], digits = 4)
```



(a) (b)

Figure 3: (a) Linear time trend for each canton, cantons with a significantly different time trend are plotted using various line types (ST1); (b) Relative incidence for Coxiellosis, 2005-2008 (ST1).

	mean	sd
Precision for region.nu	3.9539	1.7894
Precision for region.psi	61.7972	45.6382
Precision for canton.alpha (first component)	1.0040	0.3966
Precision for canton.alpha (second component)	5.3926	2.2544
Rho for canton.alpha	-0.4379	0.2282

A significantly positive linear time trend can be observed. Hence, a significant rise in disease incidence has taken place over the last four years. The summary of the obtained posterior estimates for the hyperparameters (τ_ν, τ_ψ, τ_α, τ_δ, ρ) shows, that the spatially structured regional effect can almost be neglected. Spatially structured heterogeneity is covered on a coarser resolution by the cantonal trend. The estimated correlation ρ between the cantonal and the differential trend is negative. So, the higher the cantonal intercept the less steep than the main time trend is the time trend of the respective canton. This fact can also be seen in Figure 3(a). It shows the individual time trend for each canton ($\mu + \alpha_j + (\beta_1 + \delta_j) \cdot t$). Cantons with a time trend that is significantly different from the main time trend are plotted with various line types. The two cantons with the highest disease incidence, namely Jura and Obwalden, show a significantly negative time trend, while it is positive for almost all other cantons. A plot of the mean spatial incidence of Coxiellosis for the years 2005 to 2008 is shown in Figure 3(b).

To assess the significance of the explanatory variable the output for model ST2 has to be considered. The results for the fixed effects can be called using

```
> round(ST2$summary.fixed, digits = 3)
```

	mean	sd	0.025quant	0.5quant	0.975quant	kld
(Intercept)	-9.232	0.550	-10.359	-9.216	-8.196	0.283
time.beta1	0.212	0.161	-0.091	0.207	0.545	0.036
ncalves.beta2	0.124	0.034	0.059	0.123	0.192	0.050

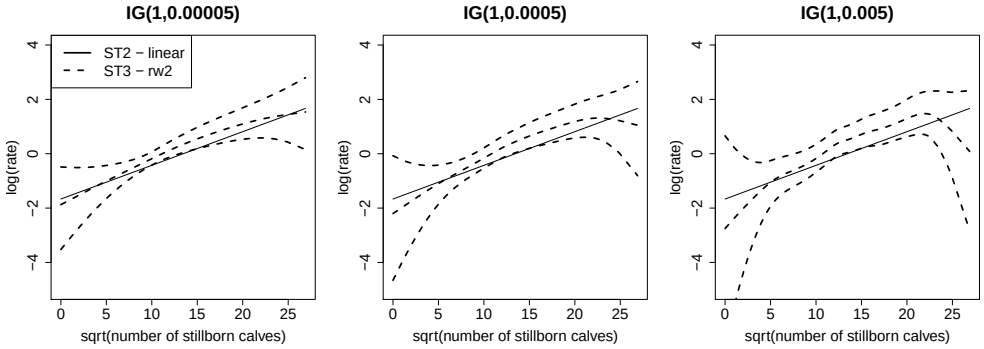


Figure 4: Estimated nonparametric trend γ for model ST3 (dashed line); additionally to the estimated posterior mean a pointwise 95%-confidence interval is plotted. The estimated linear trend (ST2) is plotted as well (solid line). The results are shown for different specifications of the prior for σ_γ^2.

The number of stillborn calves is significantly positive associated with the incidence of Coxiellosis within one region. This indicates that a real rise in disease incidence has taken place. To assure that the significance of the covariate is not confounded with the positive temporal trend, an ecological analysis has been conducted for each year separately. A significant association was found for each year, except for 2007.
The influence of the number of stillborn calves was modelled in a nonparametric fashion in models ST3 (`"rw2"`) and ST4 (`"crw2"`). Results for model ST3 using the prior suggested in Natario and Knorr-Held (2003) are shown in the first plot of Figure 4, including a 95%-confidence interval. To check, if the estimated linear effect is contained in the confidence interval of the nonparametric effect, it is also plotted. The larger (pointwise) confidence intervals for more extreme values are a typical feature of nonparametric smoothing methods. The results for `"rw2"` and `"crw2"` are almost identical, except for negligible differences in the tails of the curves.
Figure 4 also shows the results of a sensitivity analysis regarding the IGa(a,b)-hyperprior on the variance $\sigma_\gamma^2 = 1/\tau_\gamma$ for choices where the first parameter, the so-called shape parameter, is kept constant, but the prior mode $b/(a+1)$ increases from left to right. As noted by Natario and Knorr-Held (2003) the resulting curve is highly sensitive to this choice. The larger the prior mode, the more wiggly is the curve. This is the case for the `"rw2"` as well as for the `"crw2"` specification. Additionally it was found that the results barely change, if the prior mode is kept constant and only the shape parameter is varied.
Regarding model choice the same quantities as in Section 4.1 can be considered. Table 3 shows the results for the DIC and the logarithmic score. The DIC for all models including the covariate is smaller than for the model without covariate which, again, suggests a significant association between the number of stillborn calves and the Coxiellosis incidence. Not only the posterior deviance, but also the number of effective parameters is smaller when comparing models ST1 and ST2. Model ST3 provides a better fit than model ST2, but, as expected, the number of effective parameters is slightly higher. The best trade-off between model fit and complexity is offered by model ST3. The mean logarithmic score is smallest for models ST2. The PIT his-

Table 3: Summary of the posterior mean of the deviance ($\overline{D}$), the number of effective parameters (p_D) and the resulting sum, the DIC, as a measure of trade-off between model fit and complexity for models ST1, ST2, ST3 and ST4; additionally, the logarithmic score ($\overline{\text{LS}}$) is given.

Model	$\overline{D}$	p_D	DIC	$\overline{\text{LS}}$
ST1	603.7	48.6	652.3	0.455
ST2	600.5	43.5	643.9	0.448
ST3	599.2	44.1	643.3	0.449
ST4	599.7	43.9	643.6	0.449

tograms are not shown, but close to uniform; hence, all models are well calibrated.
As a general result it can be derived that a significant rise in reported cases has taken place. There is a positive association between the number of stillborn calves and the disease incidence within one region. A linear relationship might be sufficient to model this association. A drawback concerning a nonparametric formulation of the covariate is the high sensitivity towards the choice of the hyperprior.

4.3 Some comments on computer time and the accuracy of approximations

As noted in Section 2, two strategies for the exploration of the posterior marginal $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ exist, namely the GRID and the CCD strategy. Using `INLA` these strategies can be chosen using the options `int.strategy="grid"` and `"CCD"`, respectively. The CCD strategy is less precise, but takes much less computational time, as fewer support points for the integration in (1) are needed. As an example consider a spatio-temporal model like (4), which contains $h = 5$ hyperparameters (τ_v, τ_ψ, τ_α, τ_δ, ρ). Using the `INLA` default configurations for the density of the grid, the GRID strategy needs $5^h = 5^5 = 3125$ support points, while only 27 are needed for the CCD strategy. The more hyperparameters are included in the model, the larger the difference in the number of support points between the GRID and the CCD strategy. A second issue is the chosen approximation for the latent Gaussian field $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. Possible strategies are `strategy="gaussian"`, `"simplified.laplace"` and `"laplace"`, as described in Section 2.
The resulting computer time for model ST1 for all configurations on a Laptop with Intel(R) Core(TM) 2 Duo CPU T9300 2.50 GHz processor is summarized in Table 4.
The computer time needed for the GRID strategy is much higher than the time needed for the CCD strategy. The computer time increases as well when switching from the Gaussian to a simplified Laplace and a full Laplace approximation, respectively. The switch from the Gaussian to the simplified Laplace approximation takes less time than the switch to the full Laplace approximation.
The function `inla.hyperpar()` was applied to the CCD and GRID strategy results to obtained more precise approximations of the posterior marginals of single hyperparameters. The resulting CCD and GRID curves are identical for each hyperparameter. Hence, the CCD strategy is sufficient for these data. In Figure 5 the posterior marginals for all hyperparameters of model ST1 resulting from the CCD strategy are shown on log scale (except for ρ).
With regard to a comparison of the different approximation methods, the components

Table 4: First line: The computer time the R user has to wait for a result of model ST1 (in seconds); the model was run using the CCD and GRID strategy for approximation of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and all three approximation techniques for $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$. Second line: Number of observations where the computation of the predictive quantities is problematic or unreliable (in brackets).

	Gauss		SLP		FL	
	CCD	GRID	CCD	GRID	CCD	GRID
Computer time	24.73	317.45	28.65	672.74	171.45	13310.86
# of failures	164(9)	174(0)	22(1)	25(0)	0(0)	0(0)

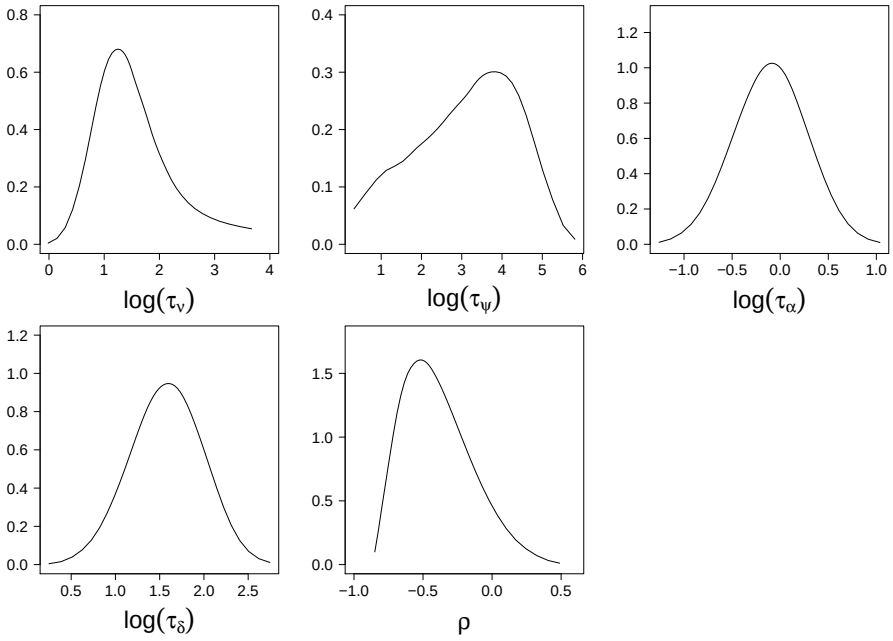


Figure 5: Posterior marginals of hyperparameters included in model ST1 on log scale (except for ρ), estimated using the CCD strategy. The results for the CCD and GRID strategy are identical.

of the latent field with the largest discrepancy between the possible approximations were determined. This was done for all random effects (i.e. $\boldsymbol{\nu}, \boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\delta}$) using the maximum symmetric Kullback-Leibler distance (SKLD) between the Gaussian and the full Laplace approximation. The resulting plots are shown in Figure 6.
For $\boldsymbol{\nu}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ a shift in location can be detected for the Gaussian approximation. The results for the simplified and the full Laplace approximation are virtually identical. Hence, the simplified Laplace approximation gives satisfactory results in terms of accuracy.
Regarding the predictive measures given by INLA the simplified Laplace approximation might not be sufficient though. As already noted in Section 3.2 and derived in Held *et al.* (2010), the approximation of the predictive measures as shown in Rue *et al.*

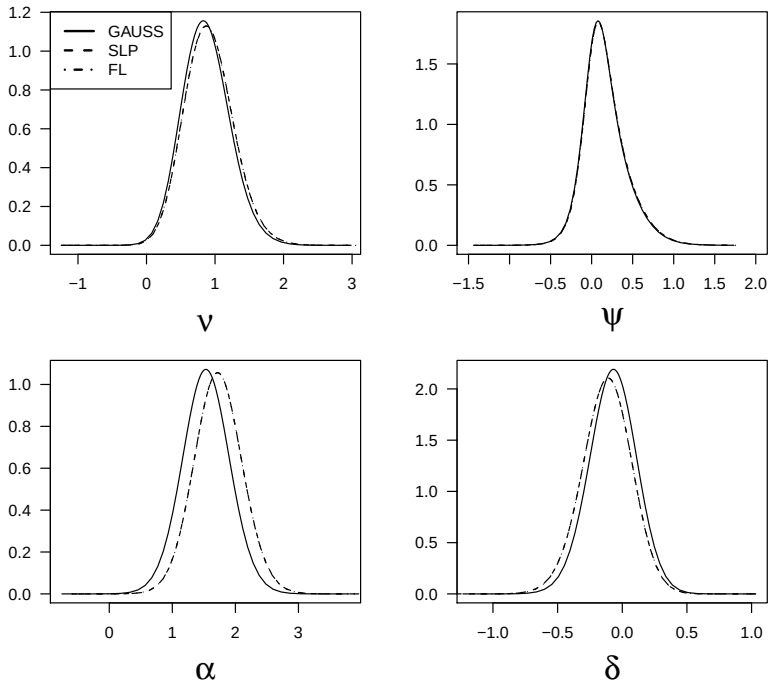


Figure 6: Posterior marginals for effects included in model ST1, estimated using the Gaussian (solid line), simplified Laplace (dashed) and full Laplace (dashed and dotted) approximation; results are shown for the marginals with the maximum symmetric Kullback-Leibler distance for the respective effect.

(2009) might fail, if the approximation of the latent field is not accurate enough. This is due to an insufficient exploration of the tail properties of involved densities. Hence, the full Laplace approximation might be obligatory to get reliable results. A feature of `INLA` is that it outputs a file which indicates the observations, where computation of the predictive measures failed. This file can be addressed using, for example,

```
> ST1$failure
```

It contains a flag for each observation. If the flag is 0, the computation of the predictive measures `cpo` and `pit` was not problematic. If it is larger than 0, there were some problems; if it is equal to 1, the obtained results are considered to be unreliable (Martino and Rue, 2009). In general, results with a `failure`-flag larger than 0 should not be used. The number of observations with problematic/unreliable results for each strategy for ST1 is shown in Table 4. The problem is mainly solved by using the full Laplace approximation. If this is not the case, the respective measures have to be computed manually by leaving out one observation in turn, re-running INLA and computing the leave-one-out predictive distribution from the respective `INLA` output (Held *et al.*, 2010).

5 Discussion

As shown within this paper, INLA can be used for Bayesian inference in spatial and spatio-temporal disease mapping models. Additionally, ecological regression can be performed involving a linear or nonparametric association between an explanatory variable and the disease incidence. The available `R` interface `INLA` can easily be handled by the user and the obtained results are useful for interpretation and suitable for model choice using the DIC or predictive measures like the logarithmic score and the PIT histogram. As INLA is a numerical approach and has a complex nature, different options for the exploration algorithm of the posterior marginals of the hyperparameters and approximation methods for the latent Gaussian field are available. This fact might make the first steps with `INLA` difficult for the standard user. As noted in Section 4.3, the default strategies give satisfactory results in this application. The computation of predictive measures often requires the use of the full Laplace approximation to obtain reliable results.

Acknowledgements

Financial support by the Swiss Federal Veterinary Office (BVET) and the Swiss National Science Foundation (SNF) is gratefully acknowledged.

References

Aitken, I. (1989). Clinical aspects and prevention of Q fever in animals, *European Journal of Epidemiology* **5**(4): 420–424.

Assunção, R., Reis, I. and Oliveira, C. (2001). Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model, *Statistics in Medicine* **20**(15): 2319–2335.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, London.

Bernardinelli, L., Clayton, D. and Montomoli, C. (1995a). Bayesian estimates of disease maps: How important are priors?, *Statistics in Medicine* **14**: 2411–2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. and Ghislandi, M. (1995b). Bayesian analysis of space-time variation in disease risk, *Statistics in Medicine* **14**: 2433–2443.

Berzuini, C. and Clayton, D. (1994). Bayesian analysis of survival on multiple time scales, *Statistics in Medicine* **13**: 823–838.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics* **43**(1): 1–59.

Clayton, D. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risk, *in* J. Cuzick et al. (eds), *Geographical and Environmental Epidemiology. Methods for Small Area Studies*, Oxford University Press, pp. 205–220.

Clayton, D., Bernardinelli, L. and Montomoli, C. (1993). Spatial correlation in ecological analysis, *International Journal of Epidemiology* **22**(6): 1193–1202.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data, *Biometrics* **65**(4): 1254–1261.

Dupuis, G., Petite, J., Péter, O. and Vouilloz, M. (1987). An important outbreak of human Q fever in a Swiss alpine valley, *International Journal of Epidemiology* **16**(2): 282–287.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society, Series C* **50**(2): 201–220.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the Amercian Statistical Association* **102**(477): 359–378.

Held, L., Schrödle, B. and Rue, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA, *in* T. Kneib and G. Tutz (eds), *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir*, Physica-Verlag, Heidelberg.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine* **19**: 2555–2567.

Kullback, S. and Leibler, R. (1951). On information and sufficiency, *The Annals of Mathematical Statistics* **22**(1): 79–86.

Lindgren, F. and Rue, H. (2008). On the second-order random walk model for irregular locations, *Scandinavian Journal of Statistics* **35**: 691–700.

Martino, S. and Rue, H. (2009). Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the `inla` program, *Technical report*, Norwegian University of Science and Technology Trondheim.

Moreno, P., Ho, P. and Vasconcelos, N. (2004). A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications, *in* S. Thrun, L. Saul and B. Schölkopf (eds), *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA.

Natario, I. and Knorr-Held, L. (2003). Non-parametric ecological regression and spatial variation, *Biometrical Journal* **45**(6): 670–688.

R Development Core Team (2005). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: `http://www.R-project.org`.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*, Chapman & Hall/CRC, London.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models, *Journal of Statistical Planning and Inference* **137**: 3177–3192.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society, Series B* **71**: 319–392.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data, *Biometrics* **60**: 1034–1042.

Schrödle, B., Held, L., Riebler, A. and Danuser, J. (2011). Using INLA for the evaluation of veterinary surveillance data from Switzerland: A case-study, *Journal of the Royal Statistical Society, Series C* **60**(2): 261–279.

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Series B* **64**(4): 583–639.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the Amercian Statistical Association* **81**(393): 82–86.

Woldehiwet, Z. (2004). Q fever (Coxiellosis): Epidemiology and pathogenesis, *Research in Veterinary Science* **77**(2): 93–100.

Wood, S. and Kohn, R. (1998). A Bayesian approach to robust binary nonparametric regression, *Journal of the American Statistical Association* **93**(441): 203–213.

Listing 1: Online Resource 1

```
#########
# Journal: Computational Statistics
# Title: A primer on disease mapping and ecological regression using INLA
# Authors: Schroedle Birgit and Leonhard Held
# Institution: University of Zurich, Switzerland
# Email: birgit.schroedle@ifspm.uzh.ch
#########

######### Online Resource 1
# additional files: cox08.txt, cox.txt, switzerland.graph

### load INLA library - for installation of INLA check www.r-inla.org
library(INLA)

### Spatial disease mapping -- Model BYM1
# read Coxiellosis data for 2008
cox.08<-read.table("cox.08.txt",header=TRUE)

# specify hyperpriors for nu and psi
prior.nu<-c(1,0.01)
prior.psi<-c(1,0.018)

# specify the formula for model BYM1 (see Section 3.2)
f.BYM1<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
 f(region.psi,model="besag",param=prior.psi,graph.file="switzerland.graph")

# run model BYM1
# strategy="laplace": use the full Laplace approximation for
# the approximation of the latent Gaussian field
# dic=1: compute the DIC
# cpo=1: compute the conditional predictive ordinate (CPO) and the
# probability integral transform (PIT)
BYM1<-inla(f.BYM1,family="poisson",E=offset,data=cox.08,
 control.inla=list(strategy="laplace"),control.compute=list(dic=1,cpo=1))

### Spatio-temporal disease mapping -- Models ST1, ST2 and ST3
# read Coxiellosis data for 2005 to 2008
cox<-read.table("cox.txt",header=TRUE)

# specify hyperpriors
prior.wishart<-c(4,1,1,0)
prior.gamma<-c(1,0.00005)

# specify and run model ST2 (see Section 3.2)
f.ST1<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
 f(region.psi,model="besag",param=prior.psi,graph.file="switzerland.graph")+
 f(canton.alpha,model="2diidwishartpart0",param=prior.wishart)+
 f(differential.delta,time,model="2diidwishartpart1")+
 time.beta1

ST1<-inla(f.ST1,family="poisson",E=offset,data=cox,
 control.inla=list(strategy="laplace"),control.compute=list(dic=1,cpo=1))

# specify and run model ST2 (see Section 3.3/3.4)
f.ST2<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
 f(region.psi,model="besag",param=prior.psi,graph.file="switzerland.graph")+
 f(canton.alpha,model="2diidwishartpart0",param=prior.wishart)+
 f(differential.delta,time,model="2diidwishartpart1")+
 time.beta1+
 ncalves.beta2
```

```
ST2<-inla(f.ST2,family="poisson",E=offset,data=cox,
 control.inla=list(strategy="laplace"), control.compute=list(dic=1,cpo=1))

# specify and run model ST3 (see Section 3.4)
f.ST3<-Y.cox~f(region.nu,model="iid",param=prior.nu)+
 f(region.psi,model="besag",param=prior.psi,graph.file="switzerland.graph")+
 f(canton.alpha,model="2diidwishartpart0",param=prior.wishart)+
 f(differential.delta,time,model="2diidwishartpart1")+
 time.beta1+
 f(ncalves.gamma,model="rw2",param=prior.gamma)

ST3<-inla(f.ST3,family="poisson",E=offset,data=cox,
 control.inla=list(strategy="simplified.laplace"),
 control.compute=list(dic=1,cpo=1))

### check INLA output --> some examples
# look at the summary of the fixed effects for model ST1
ST1$summary.fixed

# look at the summary of a random effect, e.g. region.nu
ST1$summary.random$region.nu

# look at the summary of all hyperparameters
ST1$summary.hyperpar

# look at the DIC and its components
ST1$dic

# look at all quantities needed to calculate the logarithmic score and
# a PIT histogram
ST1$cpo
ST1$pit
ST1$failure # flag: did their computation fail?

### some useful plots and functions
# plot the posterior marginal of a random effect, e.g. the first
# component of region.nu
plot(ST1$marginals.random$region.nu[[1]][,1],
 ST1$marginals.random$region.nu[[1]][,2],
 type="l",xlab="",ylab="",main=expression(nu[1]))

# plot the posterior marginal of a hyperparameter, e.g. the precision
# of nu (see summary on page 12)
plot(ST1$marginals.hyperpar[[1]][,1],ST1$marginals.hyperpar[[1]][,2],
 type="l",xlab="",ylab="",main=expression(tau[nu]))
# you can use the function inla.hyperpar(inla-object) to obtain
# more precise approximations for the marginals of single hyperparameters

# plot the result for a nonparametric explanatory variable,
# e.g. ncalves.gamma, including (pointwise) confidence bands
# be aware: the values of ncalves.gamma have be scaled using the factor
# 2.74 before the analysis, so that the mean difference between
# consecutive values is 1 (see Section 3.4)
plot(ST3$summary.random$ncalves.gamma[,1]/2.74,
 ST3$summary.random$ncalves.gamma[,2],
 type="l",xlab="",ylab="",ylim=c(-4,4),main=expression(gamma))
lines(ST3$summary.random$ncalves.gamma[,1]/2.74,
 ST3$summary.random$ncalves.gamma[,4],lty=2)
lines(ST3$summary.random$ncalves.gamma[,1]/2.74,
 ST3$summary.random$ncalves.gamma[,6],lty=2)
```

PAPER III

Spatial analysis of bluetongue cases and vaccination of Swiss cattle in 2008 and 2009

Katriina Willgert, Birgit Schrödle & Heinzpeter Schwermer

Paper published in *Geospatial Health*, 2011, **5** (2), 227-237.

Geospatial Health 5(2), 2011, pp. 227-237

Spatial analysis of bluetongue cases and vaccination of Swiss cattle in 2008 and 2009

Katriina J. E. Willgert[1], Birgit Schroedle[2], Heinzpeter Schwermer[3]

[1]Royal Veterinary College, Royal College Street, London, NW1 0TU, United Kingdom; [2]Biostatistics Division, Institute of Social and Preventive Medicine, University of Zurich, CH-8001 Zurich, Switzerland; [3]Monitoring Unit, Federal Veterinary Office, Schwarzenburgstrasse 155, CH-3003 Bern, Switzerland

Abstract. Bluetongue (BT) is a vector-borne viral disease of ruminants. The infection is widespread globally with major implications for international animal trade and production. In 2006, BT virus serotype 8 (BTV-8) was encountered in Europe for the first time, causing extensive production losses and death in susceptible livestock. Following the appearance of BTV-8 in Switzerland in 2007, a compulsory vaccination programme was launched in the subsequent year. Due to social factors and difficulties to reach animals on high pasture, the regional vaccination coverage varied across the country in both 2008 and 2009. In this study, the effect of vaccination on the spatial occurrence of BTV-8 and the associated relative disease risk in Switzerland in 2008 and 2009 were investigated by a spatial Bayesian hierarchical approach. Bayesian posterior distributions were obtained by integrated nested Laplace approximations, a promising alternative to commonly used Markov chain Monte Carlo methods. The number of observed BTV-8 outbreaks in Switzerland decreased notably from 2008 to 2009. However, only a non-significant association between vaccination coverage and the probability of a spatial unit being infected with BTV-8 was identified using the model developed for this study. The relative disease risk varied significantly across the country, with a higher relative risk of BTV-8 infection in western and north-western Switzerland where environmental conditions are more suitable for vector presence and viral transmission. Examination of the spatial correlation between disease occurrence, control measures and associated ecological factors can be valuable in the evaluation and development of disease control programmes, allowing prioritisation of areas with a high relative risk of disease.

Keywords: Bayesian hierarchical model, bluetongue, disease mapping, integrated nested Laplace approximation, vaccination,, Switzerland.

Introduction

Bluetongue (BT) is a non-contagious, arboviral disease of ruminants. The disease causes extensive economic losses in the livestock production sector and is listed as a notifiable disease by the World Organisation for Animal Health (OIE) (Mellor and Wittmann, 2002; Mintiens et al., 2008; Saegerman et al., 2008). The mortality rate can be as high as 50-70% in certain sheep breeds (Sellers, 1984; Elbers et al., 2008b) and infected countries are imposed with trade restrictions (Méroc et al., 2008). The aetiological agent of BT, bluetongue virus (BTV), belongs to the family Reoviridae and the genus *Orbivirus* (Mellor and Wittmann, 2002; Purse et al., 2005). At the present time, there are 24 recognised serotypes of BTV and a putative 25th serotype has been identified in goats (Hofmann et al., 2008; Saegerman et al., 2008). The clinical implications of BTV infection can range from asymptomatic to severe or even death. BT infection tends to be more pronounced in sheep than other domestic ruminants, such as cattle, where the disease often is sub-clinical (Sellers, 1984; Elbers et al., 2008a; Vellema, 2008).

The geographical distribution of BT is restricted to areas where there are susceptible animal hosts, competent vector species and suitable climatic conditions for viral transmission (Mellor et al., 2000; Mellor and Wittmann, 2002; Saegerman et al., 2008). BT is transmitted by biting midges of the genus *Culicoides* (Mellor and Wittmann, 2002; Purse et al., 2005). Climatic and environmental conditions, such as temperature, humidity and access to breeding sites limit the vector distribution (Sellers, 1984; Purse et al., 2005; Saegerman et al., 2008). In addition, BTV replication in the vector and host is temperature dependent and, although the vector can be active at temperatures above 10 °C, viral replication in the vector becomes prominent at temperatures above 15-18 °C (Mellor, 2000; Wittmann et al., 2002; EFSA, 2007).

Corresponding author:
Heinzpeter Schwermer
Monitoring Unit, Federal Veterinary Office
Schwarzenburgstrasse 155, CH-3003 Bern, Switzerland
Tel. +41 31 323 3053; Fax +41 31 323 9543
E-mail: heinzpeter.schwermer@bvet.admin.ch

In August 2006, BTV-8 emerged in the Netherlands, a serotype of BTV that had previously not been recorded in Europe (Elbers et al., 2008a; Mintiens et al., 2008). The infection spread rapidly across Europe in 2007 and 2008. Thus, when commercial inactivated vaccines against BTV-8 became available in 2008, several affected countries launched national vaccination campaigns in order to control the disease (Gethmann et al., 2009; Roy et al., 2009). Following the outbreak of BTV-8 in northern Europe in 2006, BT surveillance in Switzerland was intensified and increased disease awareness was encouraged among livestock keepers (Kluiters et al., 2008). For surveillance purposes, Switzerland was divided into 16 regions, hereafter referred to as BT-regions, based on animal population numbers, land area and administrative boundaries. The Principality of Liechtenstein composed a 17th BT-region (Figure 1; Kluiters et al., 2008). The first case of BTV-8 in Switzerland was detected in October 2007 (Hofmann et al., 2008). Switzerland initiated a vaccination campaign against BTV-8 in June 2008 with mandatory vaccination of cattle, sheep and goats (FVO, 2009). The vaccination campaign continued in 2009 with compulsory vaccination of cattle and sheep above 3 months of age (FVO, 2009). The vaccination programme was implemented locally by the cantonal veterinary offices and, consequently, the vaccination coverage varied regionally in both years, ranging from 24% to 82%.

The effect of vaccination on the disease prevalence in a population depends on the achieved vaccination coverage, characteristics of the disease and attributes of the targeted group (Anderson and May, 1982). Simulation models show that both the incidence and spatial spread of BTV-8 decrease with increasing vaccination coverage (Szmaragd et al., 2010a,b). By using adequate spatial models, the spatial association between disease management measures or ecological factors and disease prevalence can be explored. We applied a spatial Bayesian hierarchical approach to examine the spatial heterogeneity in relative disease risk and to investigate the association between BTV-8 occurrence and vaccination coverage in Switzerland in 2008 and 2009, while controlling for spatial variation in surveillance intensity. Since vaccination of susceptible animals reduces virus transmission and the probability of an animal acquiring infection (Szmaragd et al., 2010b), a negative correlation between vaccination coverage and the probability of a unit being infected with BTV-8 in a spatially defined area was expected.

Materials and methods

Study area and study population

Switzerland is a federal state, consisting of 20 cantons and six half cantons. The country covers an area of 41,285 km^2 and is located north of latitude 46° N in central Europe. In the south, Switzerland is divided by the Alps with a maximum altitude of 4,700 m above sea level. Out of the total land area, 37% is used for agricultural purposes. Family driven farming enterprises are most predominant, with milk being the main product of livestock keeping.

The number of cattle farms and registered cattle on

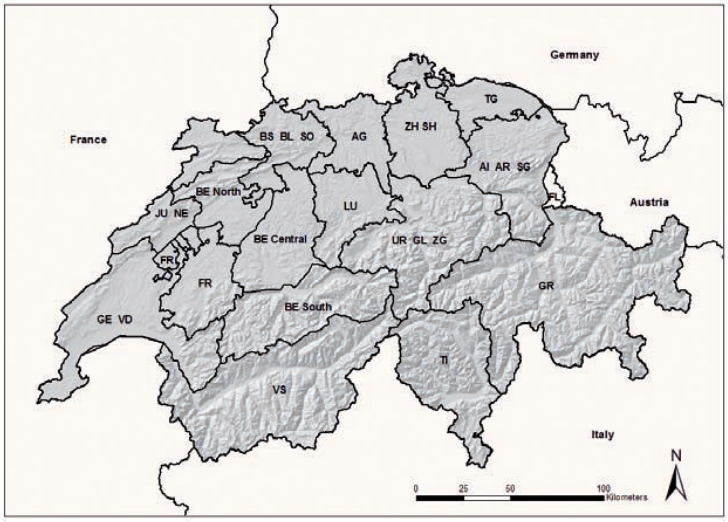


Fig. 1. Hillshade map of Switzerland showing the division of the Swiss cantons and Liechtenstein (FL) by their abbreviated names into 17 bluetongue regions.

the 1st of May in 2008 and 2009 were obtained from the Swiss Animal Movement Database (Tierverkehrsdatenbank TVD) and included in the study as the population at risk (2008: n = 1,628,435 animals/44,559 farms; 2009: n = 1,642,613 animals/44,202 farms). Since official disease control measures are applied locally by the cantonal veterinary authorities, animal and farm data were first considered at a regional level by aggregating animal data ("animal aggregated data") and farm data ("farm aggregated data") for each BT-region. Secondly, point-referenced data of each farm were analysed with explanatory variables specific to the location of the farm ("point-referenced data").

A total of 42,081 farms (90.7%) were given Cartesian coordinates based on their TVD identification number. If no geographical coordinates were available, farms were assigned the centroid coordinates of their community. ArcGIS version 9.2 (ESRI; Redlands, CA, USA) was used to map and manage spatial information. Geographical coordinates of farms were referenced in the Swiss reference system CH1903. For the regional analysis, the geographical coordinates of each farm were spatially joined to a shapefile of the 17 BT-regions in order to estimate the population at risk per BT-region. In the analysis of point-referenced data, 4,323 farms (9.3%) were excluded from the population at risk since they lacked geographical point coordinates. For farms included in the analysis of point-referenced data, the farm altitude (m above sea level) was extracted from raster elevation map DHM25 (Federal Office of Topography swisstopo) with a spatial resolution of 5 x 5 m.

Disease data

BTV-8 case data of cattle from 2008 and 2009 were extracted from the Swiss database of notifiable diseases (InfoSM) maintained by the Federal Veterinary Office (FVO). Since different surveillance activities may target separate characteristics of the population and disease (Del Rio Vilas and Pfeiffer, 2009), data from three simultaneous but independent surveillance programmes were included in the analysis: reported cases of clinical suspicion, official targeted surveillance where 10 cattle from 20 randomly selected farms of each BT-region were sampled in February to March each year (FVO, 2009), and pre-movement testing. For pre-movement tested animals, the animal identification number was linked to the TVD database in order to identify the last farm location. Data were aggregated annually since the time delay in reporting disease following infection of an animal may vary. Moreover, the time of infection of sub-clinical cases detected through the official surveillance programme and pre-movement testing is unknown. Seeing that BTV infection generally is not detected in cattle herds until 2 months after the beginning of the vector active season (Bishop et al., 2004, Conraths et al., 2009), which in Switzerland was in April in both 2008 and 2009 (FVO, 2009), cases detected through the official surveillance programme or pre-movement testing before the 1st of June were assumed to belong to the previous BT season.

An animal was considered BTV-8 positive if it had a positive polymerase chain reaction (PCR) result for BTV-8 from the national reference laboratory for BT, the Institute of Virology and Immunoprophylaxis (IVI; Mittelhäusern). A farm was regarded BTV-8 positive if at least one positive animal was detected at the farm. The diagnostic method was assumed to be 100% sensitive and specific.

Vaccination data

Yearly BTV-8 vaccination data of cattle were acquired from FVO. An animal was considered vaccinated against BTV-8 following two vaccine doses administrated within an interval of 21 to 63 days. A farm was regarded vaccinated if at least one animal at that farm had been vaccinated. Vaccine efficacy was assumed to be 100%. Any delays in reaching full protection were neglected.

Spatial modelling

Analysis of regionally aggregated data.

To estimate the relative risk of one unit being infected at a regional level, a spatial and spatio-temporal disease mapping model constructed in a Bayesian hierarchical framework was developed. Three hierarchical stages were specified as explained by Clayton and Bernardinelli (1992). The first stage describes the observed data, in this case the number of infected units per BT-region, as a function of the region-specific relative risk of disease. At stage two, the relative risk of disease for each region is modelled. The hyperpriors for all variance parameters of stage two are specified at stage three. It was assumed that the number of infected units y_i in a region i (i = 1, ..., 17) has a Poisson distribution:

$$y_i \sim \text{Poisson}\ (\lambda_i \cdot n_i) \quad (1)$$

where λ_i is the probability of one unit, either animal or farm, being BTV-8 infected in region i, and n_i the population at risk in each BT-region. The relative risk parameter λ_i, described by Besag et al. (1991), is specified using a linear predictor:

$$\eta_i = \log(\lambda_i) = \mu + \nu_i + \psi_i \quad (2)$$

where μ is the intercept. The regional term ν_i and global term ψ_i account for any unmeasured explanatory variables considered in the model (Clayton et al., 1993). The term ν_i is a spatially unstructured variable that accounts for unmeasured risk factors that vary between areas, whereas ψ_i is assumed to be structured in space and models the effect of the location, accounting for the assumption that geographically close areas are more related than distant areas (Clayton et al., 1993; Durr et al., 2005). Hence, the regional term ν_i was modelled as an independent identically (i.i.d.) normal distributed random effect, while the global term ψ_i was modelled as a so-called intrinsic Gaussian Markov random field (IGMRF) of first order on an irregular lattice (Rue and Held, 2005). Such an IGMRF accounts for spatial autocorrelation by assuming that the conditional distribution of ψ_i in region i depends on the ψ_j in neighbouring regions j. An adjacency matrix of the BT-regions was constructed, where neighbours of each region, defined by a shared common boundary, were counted and listed. To all variances, an inverse gamma hyper-prior IGa (1.0, 0.01) was assigned (Bernardinelli et al., 1995a). To account for any additional over-dispersion, the data could be modelled using a negative binomial distribution (Gschössl and Czado, 2008), which includes an additional dispersion parameter. However, the deviance information criterion (DIC) (Spiegelhalter, 2002) was higher for all models assuming a negative binomial distribution than for the corresponding Poisson models, indicating that a Poisson distribution was sufficient for this study.

Two region-specific explanatory covariates were included in the analysis to account for any spatial variation in relative disease risk: vaccination coverage (vacc_cov_i), calculated as the number of units vaccinated divided by the total number of units, and surveillance intensity (surv_int_i), estimated by dividing the number of tested units by the total number of units:

$$\eta_i = \log(\lambda_i) = \mu + \beta_1 \cdot \text{vacc\_cov}_i + \\ + \beta_2 \cdot \text{surv\_int}_i + \nu_i + \psi_i \quad (3)$$

where β denotes a fixed linear effect of the explanatory variable. The surveillance intensity in each region was included to assure that any variation in estimated disease risk was not the product of regionally differing surveillance efforts. Since the cantonal veterinary services implement legislation of animal health at a local level (Rüsch and Kihm, 2003), the surveillance intensity varies regionally and reporting of suspected disease may be biased in space and time by varying disease awareness (Kluiters et al., 2008; Del Rio Vilas and Pfeiffer, 2009).

Due to the sparsity of BTV-8 positive cases in 2009, a spatial analysis was only conducted for 2008. To assess if there was any significant change in the probability of a unit being infected between 2008 and 2009, a spatio-temporal disease mapping model was adopted. A random intercept γ_t for each year t was added to formula (3) (Knorr-Held, 2000):

$$\eta_{it} = \log(\lambda_{it}) = \mu + \beta_1 \cdot \text{vacc\_cov}_i + \\ + \beta_2 \cdot \text{surv\_int}_i + \nu_i + \psi_i + \gamma_t \quad (4)$$

This is a version of the standard non-parametric space-time model introduced by Knorr-Held (2000). Since the data considered were restricted to 2 years, a random intercept for each time point was preferred to a linear trend (Bernardinelli et al., 1995b). Due to the sparseness of BTV-8 cases, especially in 2009, any interactions between space and time were neglected. An inverse gamma hyper-prior IGa (1.0, 0.01) was assigned to the variance of γ_t.

The four mountain BT-regions Bern south (BE South), Valais (VS), Ticino (TI) and Grisons (GR) (Fig. 1) were excluded from the analysis at regional level since the risk factors associated with BT are believed to differ considerably from the rest of the country in these areas due to the distinct environmental conditions observed in the Alps (Racloz et al., 2008). In the Alps, the microclimate varies within short distances, making it difficult to form a covariate that is valid for the whole region to adjust for the local environmental conditions.

Analysis of point-referenced data.

For the analysis of point-referenced farm data, a generalised Bayesian geoadditive model was adopted (Kneib and Fahrmeir, 2006; Musio et al., 2008), where the exact location of the unit in space is known. The location of each farm in terms of easting and northing was denoted as s_{ij}. Since farms can be either BTV-8 positive or negative, it was assumed that the response y_{ij} followed a binomial distribution:

$$y_{ij} \sim \text{Binom}(\pi_{ij}, 1) \quad (5)$$

where π_{ij} denotes the probability of a farm at location s_{ij} being infected, modelled by a latent Gaussian field of the form:

$$\eta_{ij} = \text{logit}\,(\pi_{ij}) = \mu + \beta_1 \cdot \text{vacc\_cov}_{ij} + \\ + \beta_2 \cdot \text{surv\_int}_{ij} + f(s_{ij}) \quad (6)$$

where the vaccination coverage (vacc_cov_{ij}) was given the value 1 if at least one animal at the farm had been vaccinated and otherwise zero. To account for varying surveillance efforts within the study area, a 0-1-index for the surveillance intensity (surv_int_{ij}) was also included as an explanatory variable. The location effect f of each farm was modelled as a Gaussian Markov random field on a two-dimensional lattice (Rue and Held, 2005). The easting and northing of each farm were assigned to a regular grid with I rows and J columns (Besag and Kooperberg, 1995), where $I = 100$ and $J = 80$ based on the shape of Switzerland. An intrinsic second order random walk prior was adopted (Rue and Held, 2005), where the full conditional mean of a node (i,j) on the grid depends on the four direct neighbours on the grid, the four diagonal neighbours on the grid and the four direct neighbours of second order. Hence, a node (i,j) borrows strength from its 12 closest spatial neighbours. The size of the residual spatial effect is controlled by the variance, σ_f^2. An inverse gamma hyper-prior IGa (1.0, 0.01) was assigned to σ_f^2.

Due to the scarcity of BTV-8 cases, a linear spatial trend surface was also adapted with a linear predictor defined as:

$$\eta_{ij} = \text{logit}\,(\pi_{ij}) = \mu + \beta_1 \cdot \text{vacc\_cov}_{ij} + \\ + \beta_2 \cdot \text{surv\_int}_{ij} + \beta_4 \cdot i + \beta_5 \cdot j \quad (7)$$

where the indexes i and j denote the location of a farm on the spatial lattice. The conditional mean of each farm depends on the average response of the farms in the same row and the same column. This formulation gives a more global east-west and north-south effect compared to the more localized, autoregressive two-dimensional second order random walk discussed previously.

In contrast to the aggregated data, the exact altitude (altitude_{ij}) of each farm is known and can be put into model (7) as an indicator of temperature dependent dynamics of the vector population and viral replication:

$$\eta_{ij} = \text{logit}\,(\pi_{ij}) = \mu + \beta_1 \cdot \text{vacc\_cov}_{ij} + \beta_2 \cdot \text{surv\_int}_{ij} \\ + \beta_3 \cdot \text{altitude}_{ij} + \beta_4 \cdot i + \beta_5 \cdot j \quad (8)$$

Since this model is able to, at least partly, adjust for different environmental conditions in mountain areas, all data were used for this analysis.

Model choice and computation

Models were compared and selected using DIC, which takes into consideration the posterior mean deviance, a Bayesian measure of model fit, and the complexity of the model. A smaller DIC indicates a better fit of the model (Spiegelhalter, 2002).

All models were computed using integrated nested Laplace approximations (INLA), a recently proposed method for approximate Bayesian inference within latent Gaussian models. INLA has various applications, ranging from generalized mixed models to dynamic and spatio-temporal models (Rue et al., 2009; Schrödle and Held, 2010). Furthermore, INLA outperforms traditional Markov chain Monte Carlo (MCMC) methods in terms of computational time while providing very precise estimates (Rue et al., 2009). The analyses can be conducted within a powerful R environment (R Development Core Team, 2010) using the R-INLA package available at www.r-inla-org. All analyses in this study were run using the R-INLA version built on 22 October 2010.

Results

A total of 69 BTV-8 infected holdings and 131 BTV-8 infected cattle were recorded in 2008. Three affected holdings with one infected animal each were documented in 2009 (Fig. 2). Summary statistics for the 17 BT-regions are displayed in Table 1.

Parameter results of regional disease mapping models

For all regional models, the posterior distribution of the fixed effect of vaccination had a negative posterior mean (Table 2), suggesting that increasing vaccination coverage reduced the probability of a unit being infected. However, all credible intervals contained the value zero, meaning that the model did not identify a significant association between the level of regional vaccination and the probability of a unit being infected.

The number of tested units from reported cases of clinical suspicion, official targeted surveillance and pre-movement testing was used to estimate the surveillance intensity. In both the spatial and spatio-temporal regional models, the fixed effect of the surveillance intensity was not significant for the animal aggregated data, but showed a significant positive posterior mean when farm aggregated data were used

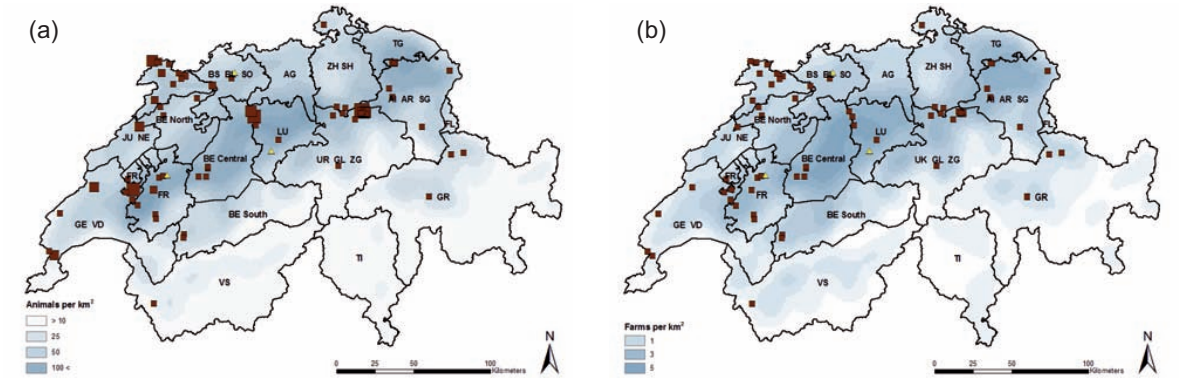


Fig. 2. Maps of Switzerland showing the spatial distribution of BTV-8 infected units recorded in 2008 (■) and 2009 (▲). Proportional symbol map of the number of BTV-8 infected cattle, ranging from 1 to 11, at each affected farm (a); and BTV-8 infected farms (b).

Table 1. Summary statistics at animal level and farm level for the 17 bluetongue regions of the population at risk (*n*), the number of detected positive BTV-8 cases (*y*) and vaccination coverage (vacc) in Switzerland in 2008 and 2009.

BT-region	Animals						Farms					
	2008			2009			2008			2009		
	n	*y*	Vacc (%)	*n*	*y*	Vacc (%)	*n*	*y*	Vacc (%)	*n*	*y*	Vacc (%)
FL	5,683	0	69	6,073	0	82	102	0	78	107	0	88
AG	91,801	0	72	91,913	0	72	2,180	0	93	2,157	0	92
AI AR SG	181,773	12	65	183,127	0	71	5,052	9	88	5,042	0	91
BS BL SO	74,302	1	73	74,557	1	77	1,828	1	94	1,803	1	93
FR	140,714	17	50	142,654	1	79	2,763	9	79	2,754	1	89
GE VD	121,871	9	52	123,545	0	78	2,575	5	79	2,551	0	88
GR	76,404	3	37	77,998	0	78	2,126	3	49	2,113	0	87
JU NE	102,520	22	68	104,757	0	78	1,766	11	91	1,775	0	92
LU	152,574	17	73	152,536	1	78	4,492	4	94	4,461	1	94
TG	74,962	0	70	75,501	0	77	1,912	0	89	1,891	0	93
TI	10,321	0	29	10,512	0	71	455	0	34	444	0	74
UR GL ZG	122,130	33	70	121,679	0	52	4,064	12	88	4,035	0	86
VS	32,327	1	32	32,826	0	67	1,537	1	50	1,521	0	82
ZH SH	110,837	4	72	111,857	0	71	2,686	4	93	2,654	0	94
BE North	79,973	6	59	80,593	0	71	2,072	4	89	2,024	0	88
BE Central	185,990	4	64	188,854	0	77	6,516	4	92	6,471	0	94
BE South	64,253	2	24	63,631	0	70	2,433	2	40	2,399	0	89

Table 2. The estimated posterior mean and the 95% posterior credible interval, given as the estimated 2.5% and 97.5% quantiles, of the vaccination coverage (β_1) and the surveillance intensity (β_2) at animal level and farm level of the spatial model (3) and spatio-temporal model (4).

	Spatial model (3)			Spatio-temporal model (4)		
	2.5%	Mean	97.5%	2.5%	Mean	97.5%
(a) Animals						
β_1	-14.59	-4.30	4.61	-4.97	-0.25	4.80
β_2	-56.05	5.79	67.51	-52.74	8.99	70.58
(b) Farm						
β_1	-8.51	-2.61	3.30	-8.00	-2.18	3.76
β_2	12.34	32.37	52.68	16.50	37.12	58.10

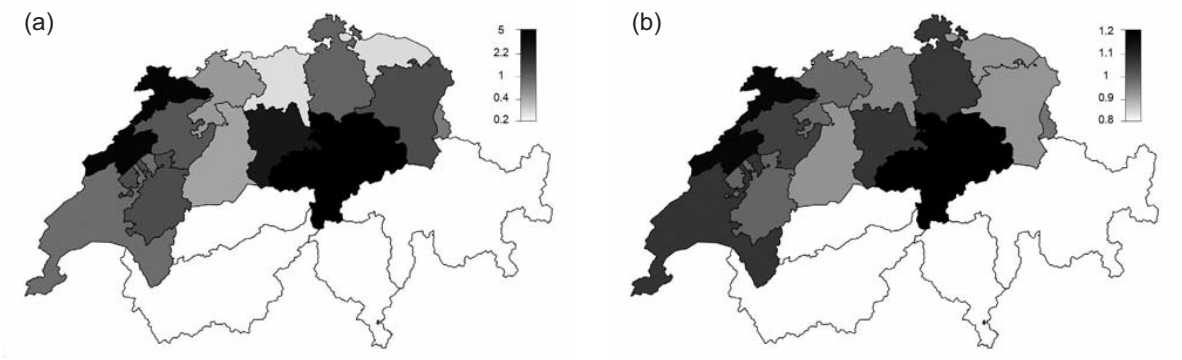


Fig. 3. The estimated relative risk of the spatial model (3) at animal level (a) and farm level (b) after accounting for regional vaccination coverage and surveillance intensity, plotted on an exponential scale ($\exp(v_i + \psi_i)$). The gray scale bar indicates the relative risk of a unit being BTV-8 infected compared to the overall risk in Switzerland in 2008. Regions in white (GR, TI, VS, BE south) were not considered in the analysis.

(Table 2), supporting that there is a positive association between surveillance intensity and the number of infected farms detected, but a less strong association between surveillance intensity and the number of infected animals detected.

The estimated random intercept for time on a log scale decreased by 3.62 for the animal aggregated data and by 1.71 for the farm aggregated data from 2008 to 2009, supporting that there was a significant change in BTV-8 prevalence from 2008 to 2009. However, when the temporal intercept was removed from model (4) (animals: DIC = 86.4; farms: DIC = 80.1) to investigate whether the decrease in prevalence of BTV-8 could be explained by the vaccination coverage alone, the DIC increased (animals: DIC = 212.7; farms: DIC = 92.5), suggesting that the improved vaccination coverage in 2009 was not sufficient to explain the temporal change in relative risk of a unit being infected.

The estimated relative risk of a unit being infected in 2008 after accounting for regional vaccination coverage and surveillance intensity is presented in Figure 3. The spatial variation between BT-regions was quite large for the models using animal aggregated data, while the range of the residual spatial risk of the farm disease mapping models was narrower, indicating that the models explain most spatial variation in relative disease risk at farm level after adjusting for regional vaccination and surveillance intensity.

Parameter results of geoadditive Bayesian models

Based on the DIC, the linear spatial trend model (7) (DIC = 501.0) showed a better prediction than model (6) assuming a localized second order random walk (DIC = 507.6), and the latter was, therefore, not considered further.

The estimated effects of the fixed coefficients of model (7) are displayed in Table 3. The posterior means of the longitude (i) and latitude (j) coefficients indicated that the probability of a farm being BTV-8 infected decreased significantly from west to east and increased significantly from south to north. Similarly to the regional analysis, the posterior mean of the vaccination coverage was non-significantly negative and the effect of surveillance intensity was significantly positive.

In a second model (8), the altitude of the farm location was added to the linear spatial trend model to

Table 3. The estimated posterior mean and 95% posterior credible interval, given as the estimated 2.5% and 97.5% quantiles, of the vaccination coverage (β_1), surveillance intensity (β_2), effect of longitude (β_4) and effect of latitude (β_5) of the geoadditive linear spatial trend model, without (model 7) and with (model 8) the effect of altitude (β_3).

	Geoadditive model (7)			Geoadditive model including altitude (8)		
	2.5%	Mean	97.5%	2.5%	Mean	97.5%
β_1	-0.99	-0.06	1.01	-1.09	-0.16	0.92
β_2	6.38	8.26	11.08	6.42	8.31	11.13
β_3	-	-	-	-1.58	-0.71	0.16
β_4	-0.04	-0.02	-0.01	-0.04	-0.02	-0.01
β_5	0.004	0.03	0.06	-0.01	0.02	0.05

account for different temperature conditions in the mountain areas. The posterior mean of the altitude was negative, supporting that the probability of a unit being BTV-8 positive decreases with increasing elevation, but the effect was not significant. Furthermore, as is illustrated in Figure 1, the altitude and latitude of Switzerland are highly correlated. As a consequence, the effect of the latitude decreased, indicating that the higher altitude in the Alps accounts for parts of the decreased probability of a farm being BTV-8 infected in the south of Switzerland. The negative effect of the vaccination coverage became slightly stronger when adjusting for the farm altitude, while the effect of the surveillance intensity remained similar to model (7). The DIC was lower when the farm altitude was included as an explanatory variable (DIC = 500.5) but the difference was negligible and, thus, there was no strong preference for either of the models. The residual disease risk after accounting for all explanatory variables decreases from north-west to south-east (Fig. 4). Hence, a trend in the spatial risk surface could still be found after adjustment for the altitude of the farms.

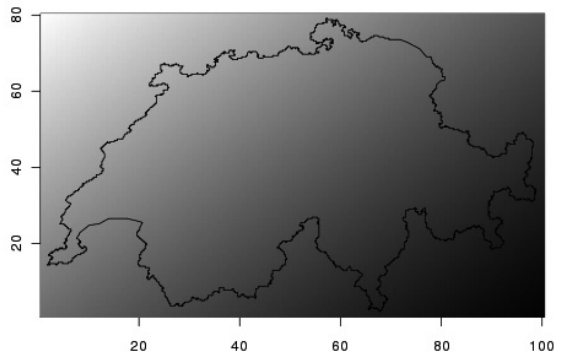


Fig. 4. The estimated relative linear spatial trend obtained from the geoadditive linear spatial trend model (8), adjusted for vaccination coverage, surveillance intensity and altitude. The relative risk ranges from 0.1 (black) to 6.0 (white). The boundary of Switzerland is plotted as an indication of the area under observation.

Discussion

In this study, the effect of vaccination on the regional and local occurrence of BTV-8 in Switzerland was investigated. Vaccination against BTV-8 was initiated in 2008, followed by a pronounced decrease in registered BTV-8 cases in 2009. The probability of a unit being infected with BTV-8 was expected to be inversely correlated with vaccination coverage. Although the results of the study point in this direction, a significant effect of vaccination coverage on disease occurrence was not demonstrated.

It may seem contradictory that vaccination is a poor explanatory variable of the probability of a unit being BTV-8 infected. There is an established association between vaccination and the occurrence and spread of infection, and vaccination is an important tool in the control of BT (Caporale et al., 2004; EC, 2009). Nevertheless, in the Alps, there were large areas with a low vaccination coverage as well as a low prevalence of BTV-8. Within the geoadditive model, the effect of vaccination coverage became stronger when the farm altitude was accounted for, suggesting that other factors than solely vaccination influence the probability of a unit being infected with BTV-8. For instance, Szmaragd et al. (2010b) showed that both the vaccination coverage and temperature significantly affect the range of an outbreak.

Additionally, the government prioritised vaccination of areas with a history of BTV-8 outbreaks. Such spatially targeted vaccination could obscure any existing correlation between vaccination coverage and disease prevalence since the vaccination coverage becomes higher in areas with an elevated underlying disease risk. It is also possible that the vaccine uptake increased with proximity to the affected area as presumed infection risk may affect the motivation of animal keepers to vaccinate their livestock (Elbers et al., 2010).

As the analysis was limited to annual time periods, different levels of population susceptibility and disease occurrence throughout the year were not considered. In 2008, the vaccination campaign was not initiated until June and the vaccination of many animals was delayed since they were on summer pasture in the Alps (EC, 2009; FVO, 2009). Cattle receive two vaccine doses 3 to 9 weeks apart, followed by an estimated delay of 3 weeks to reach full protection (Gethmann et al., 2009). Thus, animals considered vaccinated may not have attained full protection against BTV-8 until later in the vector active season. In regions of Italy where mass vaccination against BTV-2 and BTV-9 was not achieved before the start of a new epidemic, vaccination did not significantly reduce disease occurrence, while the spread of infection and disease occurrence decreased significantly in regions with a high vaccination coverage before the start of a new epidemic (Caporale et al., 2004).

Disease outbreaks may appear clustered as a consequence of spatial heterogeneity in surveillance efforts (Kluiters et al., 2008; Birch et al., 2009; Del Rio Vilas

and Pfeiffer, 2009). To reduce surveillance bias, surveillance intensity was included in the model as an explanatory variable. At farm level, case detection was clearly correlated with surveillance intensity, supporting that spatial heterogeneity in surveillance intensity needs to be considered when assessing spatial occurrence of disease at a national level (Kluiters et al., 2008). The model did not find any significant association between the detection of BTV-8 infected cattle and surveillance intensity, possibly due to a lower surveillance intensity at animal level than farm level.

After accounting for vaccination coverage and surveillance intensity, the probability of a farm being infected with BTV-8 increased significantly from east to west and south to north of Switzerland. The observed spatial effect could be related to several ecological factors, such as vector presence and activity, temperature, host density and the disease situation in neighbouring countries. Local environmental conditions and microclimate affect the presence of midges, and the abundance of *Culicoides* midges varies significantly between different regions of Switzerland (Cagienard et al., 2006; Casati et al., 2009). Due to the environmental conditions, north and northwest of Switzerland is more suitable for *Culicoides* habitation (Racloz et al., 2007, 2008). The lower relative risk of BTV-8 infection in the south of Switzerland was explained, at least partly, by the higher altitude in the mountain area in the south, and the effect of the latitude became non-significant when the altitude was adjusted for in the model. This agrees with previous studies where lower BTV prevalence was observed at higher elevations in otherwise endemic zones (Ward and Carpenter, 1996) and could be related to a shorter vector active season at higher altitudes (EFSA, 2007) as well as temperature dependent characteristics of vector behaviour and viral replication (Ward and Carpenter, 1996). The northwest of Switzerland is also the most livestock dense area. Simulation studies of BTV spread in Scotland (Szmaragd et al., 2010a) showed that the spatial risk of a holding being BTV-8 infected was associated with livestock and farm densities, where areas of high animal densities allow for onward transmission whereas spread of infection is limited in regions with lower livestock densities.

The maps of residual spatial variation are representations of the relative risk of a unit being BTV-8 infected in a defined area relative to the overall risk in Switzerland (Fig. 3). Nevertheless, Bayesian hierarchical models assume a constant spatial and temporal risk within each region and timeframe. However, the spatial and spatio-temporal divisions may not coincide with changes in disease risk (Pascutto et al., 2000). Since BT is a vector-borne, transboundary disease, BTV-8 prevalence and achieved vaccination coverage in neighbouring countries may also influence the relative disease risk in Switzerland. In 2008, the French vaccination coverage for BTV-8 was low (36%), which led to further spread of the disease (EC, 2009) and could contribute to a higher relative disease risk in the west of Switzerland. In Germany, on the other hand, a national vaccination coverage of 85% was attained with a substantial reduction in disease cases (EC, 2009). Nevertheless, these are national figures and may not be representative for areas bordering Switzerland. In Austrian regions adjacent to Switzerland, the vaccination coverage was also high, and in Italy, BTV-8 infection was secluded to a limited area with a vaccination coverage considered to be satisfactory (EC, 2009), reducing any potential spread of BTV-8 into Switzerland from the east or south.

The role of susceptible species apart from cattle in the maintenance and spread of BTV-8 was not considered. The level of infection in other susceptible populations and vaccination coverage in sheep and goats could affect the relative risk of BTV-8 infection observed in cattle. For instance, BTV-8 has been recorded in two chamois and one roe deer in Switzerland and wildlife could act as a potential reservoir for BTV-8 (Szmaragd et al., 2010a). Thus, vaccination data and BTV-8 prevalence in other ruminants should ideally be included in future studies.

Conclusion

This paper has outlined the application of spatial Bayesian hierarchical models to assess the relative disease risk of BTV-8 at a national level and the effect of vaccination as a method of control of BTV-8. By combining data from different surveillance paths with ecological factors, estimates of relative disease risk were obtained, allowing disease control efforts to be targeted to high risk areas. This is especially a priority in countries where resources available for disease control are limited. Examination of the spatial correlation between disease occurrence, control measures and associated ecological factors is valuable in the evaluation and development of disease control programmes. Especially for diseases with several plausible control strategies, the method could be adapted to evaluate and compare the effect of alternative control measures and serve as a basis in policy development.

Acknowledgements

The authors thank the staff at the Swiss Federal Veterinary Office for their support. The study was funded by the Swiss Federal Veterinary Office. RVC manuscript ID number: P/MSc/000108.

References

Anderson RM, May RM, 1982. Directly transmitted infectious diseases: control by vaccination. Science 215, 1053-1060.

Bernardinelli L, Clayton D, Montomoli C, 1995a. Bayesian estimates of disease maps: how important are priors? Stat Med 14, 2411-2431.

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, 1995b. Bayesian analysis of space-time variation in disease risk. Stat Med 14, 2433-2443.

Besag J, Kooperberg C, 1995. On conditional and intrinsic autoregression. Biometrika 82, 733-746.

Besag J, York J, Mollié A, 1991. Bayesian image restoration with two applications in spatial statistics. Ann Inst Stat Math 43, 1-59.

Birch CPD, Chikukwa AC, Hyder K, Del Rio Vilas VJ, 2009. Spatial distribution of the active surveillance of sheep scrapie in Great Britain: an exploratory analysis. BMC Vet Res 5, 23.

Bishop AL, Spohr LJ, Barchia IM, 2004. Effects of altitude, distance and waves of movement on the dispersal in Australia of the arbovirus vector, *Culicoides brevitarsis* Kieffer (Diptera: Ceratopogonidae). Prev Vet Med 65, 135-145.

Cagienard A, Griot C, Mellor PS, Dension E, Stärk KDC, 2006. Bluetongue vector species of *Culicoides* in Switzerland. Med Vet Entomol 20, 239-247.

Caporale V, Giovannini A, Patta C, Calistri P, Nannini D, Santucci U, 2004. Vaccination in the control strategy of bluetongue in Italy. Dev Biol (Basel) 119, 113-127.

Casati S, Racloz V, Delécolle JC, Kuhn M, Mathis A, Griot C, Stärk KDC, Vanzetti T, 2009. An investigation on the *Culicoides* species composition at seven sites in southern Switzerland. Med Vet Entomol 23, 93-98.

Clayton D, Bernardinelli L, 1992. Bayesian methods for mapping disease risk. In: Elliot P, Cuzick J, English D, Stern R (eds.), Geographical and Environmental Epidemiology. Methods for Small Area Studies, Oxford University Press, 205-220 pp.

Clayton DG, Bernardinelli L, Montomoli C, 1993. Spatial correlation in ecological analysis. Int J Epidemiol 22, 1193-1202.

Conraths FJ, Gethmann JM, Staubach C, Mettenleiter TC, Beer M, Hoffmann B, 2009. Epidemiology of bluetongue virus serotype 8, Germany. Emerg Infect Dis 15, 433-435.

Del Rio Vilas VJ, Pfeiffer DU, 2009. The evaluation of bias in scrapie surveillance: a review. Vet J 185, 259-264.

Durr PA, Tait N, Lawson AB, 2005. Bayesian hierarchical modelling to enhance the epidemiological value of abattoir surveys for bovine fasciolosis. Prev Vet Med 71, 157-172.

EC, 2009. Emergency vaccination against bluetongue in the EU Nov 2007-Dec 2008, evaluation report. European Commission SANCO/7045/2009.

EFSA, 2007. Bluetongue vectors and vaccines. The EFSA Journal 479, 1-29.

Elbers ARW, Backx A, Meroc E, Gerbier G, Staubach C, Hendrichx G, van der Spek A, Mintiens K, 2008a. Field observations during the the bluetongue serotype 8 epidemic in 2006 I. Detection of first outbreaks and clinical signs in sheep and cattle in Belgium, France and the Netherlands. Prev Vet Med 87, 21-30.

Elbers ARW, Backx A, Mintiens K, Gerbier G, Staubach C, Hendrickx G, van der Spek A, 2008b. Field observations during the bluetongue serotype 8 epidemic in 2006 II. Morbidity and mortality rate, case fatality and clinical recovery in sheep and cattle in the Netherlands. Prev Vet Med 87, 31-40.

Elbers ARW, de Koeijer AA, Scolamacchia F, van Rijn PA, 2010. Questionnaire survey about the motives of commercial livestock farmers and hobby holders to vaccinate their animals against bluetongue virus serotype 8 in 2008-2009 in the Netherlands. Vaccine 28, 2473-2481.

FVO, 2009. Blauzungenkrankheit in der Schweiz. Federal Veterinary Office. Available at: <http://www.bvet.admin.ch/gesundheit_tiere/01973/02437/index.html?lang=de&download=NHzLpZeg7t,lnp6I0NTU042l2Z6ln1acy4Zn4Z2qZpnO2Yuq2Z6gpJCEeYJ_fWym162epYbg2c_JjKbNoKSn6A—; accessed 5 May 2010.

Gethmann J, Hüttner K, Heyne H, Probst C, Ziller M, Beer M, Hoffmann B, Mettenleiter TC, Conraths FJ, 2009. Comparative safety study of three inactivated BTV-8 vaccines in sheep and cattle under field conditions. Vaccine 27, 4118-4126.

Gschössl S, Czado C, 2008. Modelling count data with overdispersion and spatial effects. Stat Pap 49, 531-552.

Hofmann MA, Griot C, Chaignat V, Perler L, Thür B, 2008. Bluetongue disease reaches Switzerland. Schweiz Arch Tierheilkd 150, 49-56.

Hofmann MA, Renzullo S, Mader M, Chaignat V, Worwa G, Thuer B, 2008. Genetic characterization of Toggenburg orbivirus, a new bluetongue virus, from goats, Switzerland. Emerg Infect Dis 14, 1855-1861.

Kluiters G, Chaignat V, Schwermer H, 2008. Spatial distribution of bluetongue surveillance and cases in Switzerland. Schweiz Arch Tierheilkd 150, 543-522.

Kneib T, Fahrmeir L, 2006. Structured additive regression for categorical space-time data: a mixed model approach. Biometrics 62, 109-118.

Knorr-Held L, 2000. Bayesian modelling of inseperable space-time variation in disease risk. Stat Med 19, 2555-2567.

Mellor PS, 2000. Replication of arboviruses in insect vectors. J

Comp Pathol 123, 231-247.

Mellor PS, Boorman J, Baylis M, 2000. *Culicoides* biting midges: their role as arbovirus vectors. Annu Rev Entomol 45, 307-340.

Mellor PS, Wittmann EJ, 2002. Bluetongue virus in the Mediterranean basin 1998-2001. Vet J 164, 20-37.

Méroc E, Faes C, Herr C, Staubach C, Verheyden B, Vanbinst T, Vandenbussche F, Hooyberghs J, Aerts M, De Clercq K, Mintiens K, 2008. Establishing the spread of bluetongue virus at the end of the 2006 epidemic in Belgium. Vet Microbiol 131, 133-144.

Mintiens K, Méroc E, Mellor PS, Staubach C, Gerbier G, Elbers ARW, Hendrickx G, De Clercq K, 2008. Possible routes of introduction of bluetongue virus serotype 8 into the epicentre of the 2006 epidemic in north-western Europe. Prev Vet Med 87, 131-144.

Musio M, Augustin H, von Wilpert K, 2008. Geoadditive Bayesian models for forestry defoliation data: a case study. Environmetrics 19, 630-642.

Pascutto C, Wakefield JC, Best NG, Richardson S, Bernardinelli L, Staines A, Elliott P, 2000. Statistical issues in the analysis of disease mapping data. Stat Med 19, 2493-2519.

Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PPC, Baylis M, 2005. Climate change and the recent emergence of bluetongue in Europe. Nat Rev Microbiol 3, 171-181.

R Development Core Team, 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at: http://www.R-project.org (accessed 9 June 2010).

Racloz V, Presi P, Vounatsou P, Schwermer H, Casati S, Vanzetti T, Griot C, Stärk KDC, 2007. Use of mapping and statistical modelling for the prediction of bluetongue occurrence in Switzerland based on vector biology. Vet Ital 43, 513-518.

Racloz V, Venter G, Griot C, Stärk KDC, 2008. Estimating the temporal and spatial risk of bluetongue related to the incursion of infected vectors into Switzerland. BMC Vet Res 4, 42.

Roy P, Boyce M, Noad R, 2009. Prospects for improved bluetongue vaccines. Nat Rev Microbiol 7, 120-128.

Rue H, Held L, 2005. Gaussian Markov random fields, Chapman and Hall, 101-117 pp.

Rue H, Martino S, Chopin N, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Series B Stat Methodol 71, 319-392.

Rüsch P, Kihm U, 2003. The federal system of veterinary services in Switzerland. Rev Sci Tech 22, 423-432.

Saegerman C, Berkvens D, Mellor PS, 2008. Bluetongue epidemiology in the European Union. Emerg Infect Dis 14, 539-544.

Schrödle B, Held, L, 2010. A primer on disease mapping and ecological regression using INLA. Comput Stat 26, 241-258.

Sellers RF, 1984. Bluetongue in Africa, the Mediterranean region and Near East - disease, virus and vectors. Prev Vet Med 2, 371-378.

Spiegelhalter DJ, 2002. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 64, 583-639.

Szmaragd C, Gunn GH, Gubbins S, 2010a. Assessing the consequences of an incursion of a vector-borne disease. II. Spread of bluetongue in Scotland and impact of vaccination. Epidemics 2, 139-147.

Szmaragd C, Wilson AJ, Carpenter S, Wood JLN, Mellor PS, Gubbins S, 2010b. The spread of bluetongue virus serotype 8 in Great Britain and its control by vaccination. PLoS One 5, e9353.

Vellema P, 2008. Bluetongue in sheep: question marks on bluetongue virus serotype 8 in Europe. Small Rumin Res 76, 141-148.

Ward MP, Carpenter TE, 1996. Simulation modeling of the effect of climatic factors on bluetongue virus infection in Australian cattle herds. I. Model formulation, verification and validation. Prev Vet Med 27, 1-12.

Wittmann EJ, Mellor PS, Baylis M, 2002. Effect of temperature on the transmission of orbiviruses by the biting midge, *Culicoides sonorensis*. Med Vet Entomol 16, 147-156.

PAPER IV

Spatio-temporal disease mapping using INLA

Birgit Schrödle & Leonhard Held

Paper published in *Environmetrics*, 2010, DOI: 10.1002/env.1065.

Special Issue Paper **Environmetrics**

Received: 30 November 2009, Revised: 14 May 2010, Accepted: 3 August 2010, Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI: 10.1002/env.1065

Spatio-temporal disease mapping using INLA†

Birgit Schrödle[a]* and Leonhard Held[a]

Spatio-temporal disease mapping models are a popular tool to describe the pattern of disease counts. They are usually formulated in a hierarchical Bayesian framework with latent Gaussian model. So far, computationally expensive Markov chain Monte Carlo algorithms have been used for parameter estimation which might induce a large Monte Carlo error. An alternative method using integrated nested Laplace approximations (INLA) has recently been proposed. A major advantage of INLA is that it returns accurate parameter estimates in short computational time. Additionally, the deviance information criterion is provided for Bayesian model choice. This paper describes how several parametric and nonparametric models and extensions thereof can be fitted to space–time count data using INLA. Particular emphasis is given to the appropriate choice of linear constraints to ensure identifiability of the parameter estimates. The models are applied to counts of Salmonellosis in cattle reported to the Swiss Federal Veterinary Office 1991–2008. Copyright © 2010 John Wiley & Sons, Ltd.

key words: disease mapping; INLA; linear constraints; spatio-temporal models; space–time interaction

1. INTRODUCTION

Spatio-temporal disease mapping models are a popular tool to describe the pattern of disease counts and to identify regions with unusual incidence levels, time trends or both. This class of models is usually formulated within a hierarchical Bayesian framework with latent Gaussian model (Besag *et al.*, 1991; Banerjee *et al.*, 2004). Several proposals have been made including a parametric (Bernardinelli *et al.*, 1995b; Assunção *et al.*, 2001) and nonparametric (Knorr-Held, 2000; Lagazio *et al.*, 2003; Schmid and Held, 2004) formulation of the time trend and the respective space–time interactions. To obtain the respective parameter estimates computationally expensive Markov chain Monte Carlo (MCMC) algorithms are typically used, which might induce a large Monte Carlo error of the parameter estimates. Furthermore, if complex spatial and spatio-temporal models are to be fitted, specific block-sampling algorithms have to be applied to get reliable estimates (Knorr-Held and Rue, 2002; Schmid and Held, 2004).

Recently, an approximate method for parameter estimation in latent Gaussian models was proposed in Rue *et al.* (2009). This method uses integrated nested Laplace approximations (INLA) to compute the posterior marginals of all parameters of interest. A major advantage of INLA is that it returns precise parameter estimates in short computational time and is easy to use. Additionally, the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) is provided by INLA for Bayesian model choice.

This paper describes how several parametric and nonparametric models (Bernardinelli *et al.*, 1995b; Knorr-Held, 2000) and extensions thereof can be fitted to space–time count data using INLA. We give particular emphasis on the appropriate choice of linear constraints to ensure identifiability of the parameter estimates. As far as we know, this is the first general theoretically sound discussion of the incorporation of linear constraints in space–time disease mapping models, in contrast to ad-hoc "on-the-fly" centering of MCMC samples. The different models will be applied to counts of Salmonellosis in cattle reported to the Swiss Federal Veterinary Office between 1991 and 2008, see Table 1.

Salmonellosis in cattle is a widespread disease in Switzerland. Disease counts are reported in all 184 regions of Switzerland. Additionally, data from the Principality of Liechtenstein are also available. For each region and each year the number of infected herds and the number of herds is known. An infected herd has at least one diseased animals. To adjust for the different number of herds per region the number of herds is included in the model as an offset.

This paper is organized as follows: Section 2 gives an introduction to INLA and describes the DIC used for model choice. Section 3 gives an overview of spatio-temporal disease mapping models in a parametric and nonparametric setting. Results of model choice and parameter estimation for the Salmonellosis data are presented in Section 4. We close with some discussion in Section 5.

* *Correspondence to: Birgit Schrödle, University of Zurich, Hirschengraben 84, Zurich, Switzerland. E-mail: birgit.schroedle@ifspm.uzh.ch*

a *Birgit Schrödle, Leonhard Held*
Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Zurich, Switzerland

†This paper is published in Environmetrics as a special issue TIES-GRASPA, edited by Editors M. Scott and D. Cocchhi,University of Zurich, Hirschengraben 84, Zurich, Switzerland.

Table 1. Number of cattle herds infected with Salmonellosis, 1991–2008

91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
69	73	30	48	61	83	99	64	79	56	42	52	39	32	16	22	27	17

2. APPROXIMATE BAYESIAN INFERENCE USING INLA

Integrated nested Laplace approximations (INLA) are a recently proposed method (Rue *et al.*, 2009) for approximate Bayesian inference in structured additive regression models with latent Gaussian field. The methodology is particularly attractive, if the latent Gaussian model is a Gaussian Markov random field (GMRF) (Rue and Held, 2005) with precision matrix $\boldsymbol{Q}$ controlled by a few hyperparameters $\boldsymbol{\theta}$. For such models an analytical computation of the posterior marginals of the unknown parameters is not available. Therefore, the standard solution to obtain estimates are MCMC methods, but they are not without problems: computational time is long, parameter samples can be highly correlated and estimates may have a large Monte Carlo error. Application of MCMC to space–time models is particularly difficult since often a strong posterior dependence between components of the latent spatial or spatio-temporal field is present.

In contrast, INLA provides very accurate approximations to the posterior marginals in relatively short computational time. In the following we explain briefly how INLA computes posterior marginal distributions of parameters of interest, for details see Rue *et al.* (2009).

Let $\boldsymbol{x}$ denote the vector of all Gaussian variables and $\boldsymbol{\theta}$ the vector of hyperparameters, which are not necessarily Gaussian. Of main interest is typically the marginal posterior density

$$\pi(x_i|\boldsymbol{y}) = \int_{\theta}\int_{\boldsymbol{x}_{-i}} \pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x}_{-i}\, \mathrm{d}\boldsymbol{\theta}$$

given some data $\boldsymbol{y}$ for each component x_i of the Gaussian random vector $\boldsymbol{x}$. Here, $\boldsymbol{x}_{-i}$ denotes $\boldsymbol{x}$ with the ith component omitted. The marginal posterior density of x_i can also be written as

$$\pi(x_i|\boldsymbol{y}) = \int_{\theta} \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\, \pi(\boldsymbol{\theta}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta}. \tag{1}$$

The key feature of the INLA approach is to construct a nested approximation of Equation (1). For performing this approximation it is helpful that the precision matrix $\boldsymbol{Q}$ of the Gaussian field is sparse, if the field has the Markov property. Hence, numerical methods for sparse matrices can be used, which are much quicker than general algorithms for dense matrices (Rue and Held, 2005).

The second component in the integral (1), the marginal posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of the hyperparameters $\boldsymbol{\theta}$, can be approximated using

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{2}$$

which is basically the Laplace approximation as described in Tierney and Kadane (1986). In Equation (2), $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ denotes the Gaussian approximation (Rue and Held, 2005) to the full conditional distribution of $\boldsymbol{x}$ and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\boldsymbol{x}$ for a given $\boldsymbol{\theta}$. The main use of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is to integrate out the uncertainty with respect to $\boldsymbol{\theta}$ in Equation (1) numerically. It is important to find good support points $\boldsymbol{\theta}_k$ for a numerical integration of Equation (1).

Rue *et al.* (2009) have proposed three approaches to approximate the first component $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ of the integral in Equation (1): a Gaussian approximation, a full Laplace approximation and a simplified Laplace approximation. Each approach has different features and the results are supposed to be differently accurate. The simplest approximation $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ to $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation. According to Rue and Martino (2007), this method often gives quite satisfactory results in short computation time. However, there can be numerical errors in the location and/or errors due to the lack of skewness of the Gaussian approximation. It can be improved through applying another Laplace approximation to $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. This "full Laplace" approximation is supposed to be most accurate, but there is an alternative, called "simplified Laplace" approximation, which is less expensive from a computational point of view with only a slight loss of accuracy. This method is based on a series expansion of the full Laplace approximation. Unfortunately, the computation time for the simplified or the full Laplace approximation will increase considerably, if a large number of linear constraints on $\boldsymbol{x}$ has to be incorporated. See Section 3.3 for a discussion of such constraints with regard to nonparametric space-time models.

As an approximation of the posterior marginal density (1) we obtain

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\boldsymbol{\theta}_k, \boldsymbol{y})\, \widetilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y})\Delta_k. \tag{3}$$

For substitution of the integral in Equation (1) an area weight Δ_k has to be assigned to each $\boldsymbol{\theta}_k$. Its size depends on the actual strategy of choosing appropriate $\boldsymbol{\theta}_k$'s.

According to Rue *et al.* (2009), INLA returns accurate parameter estimates for a wide range of models. This was shown by a series of case studies confirming the agreement between INLA and MCMC. More recently, a variety of papers dealing with the application and accuracy of INLA in very different settings was published. As an example, Martino *et al.* (2010) show how INLA can be used for inference in survival models. Paul *et al.* (2010) discuss its advantages with regard to a bivariate meta-analysis of diagnostic test studies. In Schrödle *et al.* (2010) a comparison of INLA and MCMC for a spatio-temporal setting similar to (6) concerning accuracy of the parameter estimates is presented. The INLA approximations of the marginal posterior densities and the corresponding MCMC histograms are virtually identical, if the simplified or the full Laplace approximation is used.

An attractive feature of the INLA approach is that the deviance information criterion (DIC) can be computed. Using the DIC models in a Bayesian framework can be compared with respect to fit and complexity. It combines a measure for model fit, namely the posterior mean of the deviance $\bar{D}$, with a measure for model complexity, the number of effective parameters p_D. The quantity p_D is defined as the posterior mean of the deviance minus the deviance at the posterior means of the parameters of interest. According to Spiegelhalter *et al.* (2002) the DIC is the sum of both

$$\text{DIC} = \bar{D} + p_D.$$

The model with the lowest DIC provides the best trade-off between fit and model complexity. Both components are computed in INLA if the option `dic=1` is set, for details see Rue *et al.* (2009,Chapter 6.4).

By INLA, no samples of the posterior marginals are drawn, since they are approximated directly. Hence, more elaborate tools for model choice like, e.g., the Dawid–Sebastiani-Score and the ranked probability score (Gneiting and Raftery, 2007) cannot be computed from the INLA output. On the contrary, other measures for model choice and calibration can be obtained by INLA in a simpler way than using MCMC. In Held *et al.* (2010) it is described how easily the cross-validated logarithmic score and the probability integral transform (Czado *et al.*, 2009) can be obtained without re-running the model. Note that the marginal likelihood, an alternative model choice criterion, is also available from INLA but cannot be used in our setting since we use improper priors (see Section 3.5). Simultaneous credible regions are currently not available in INLA, but can be computed from the MCMC samples (Held, 2004).

From a computational point of view INLA can be used almost as a black box. A program called `inla` written in the language `C` is available which is built on the `GMRFlib` library already on-hand as supplement for Rue and Held (2005). It is open source and can be freely downloaded from the webpage http://www.math.ntnu.no/∼hrue/GMRFLib/ for Windows, Macintosh and Linux. Additionally, an interface to `R` (R Development Core Team, 2005) is available on http://www.r-inla.org as a package called `INLA`. Using `INLA` model specification and processing of the results can be done directly in `R`. Details on the usage of `inla` and `INLA` can be found in Martino and Rue (2009). Some issues of the specification of spatio-temporal models are discussed in Section 3.4. All analyses within this paper were run using the `INLA` package build on the 24th of November 2009 on `inla` version 1.624.

3. SPACE–TIME MODELING

Many models describing space and time variation of disease risk have been proposed in the literature. Most of them are based on the hierarchical Bayesian model introduced in Besag *et al.* (1991). Waller *et al.* (1997) applied this model to each time point separately which implies that the estimated spatial pattern can be completely different at each time point. Bernardinelli *et al.* (1995b) formulated a model with spatial main effects for the whole time period including a linear time trend. This was modified by Assunção *et al.* (2001) using a second degree polynomial. Knorr-Held (2000) gave a comprehensive description of possible interactions between time and space in a nonparametric setting. This approach was applied to a number of different models, e.g. to age–period–cohort models (Lagazio *et al.*, 2003). Recently, an alternative approach has been proposed by Martínez-Beneito *et al.* (2008) that offers an autoregressive approach to disease mapping by adapting ideas from autoregressive time series and spatial modeling to link information in time and space, respectively. Within this paper the proposals made by Bernardinelli *et al.* (1995b) and Knorr-Held (2000) and extensions thereof will be applied.

The following section gives an introduction to models for spatio-temporal disease mapping. Sections 3.1 and 3.2 describe parametric (Bernardinelli *et al.*, 1995b) and nonparametric (Knorr-Held, 2000) approaches, respectively, for modeling the time trend in the different regions. The framework in Knorr-Held (2000) is extended by adopting a random walk of second order for the main temporal trend and the corresponding interaction effects. In Section 3.3 a general approach to specify and incorporate linear constraints on nonparametric space–time interaction terms is proposed. Model specification using `INLA` and the choice of priors will also be discussed.

3.1. Models with parametric time trend

Let y_{it} denote the number of infected herds in area i and year t and let m_{it} denote the corresponding number of herds at risk. All models assume a Poisson observation model for y_{it} with disease intensity $\lambda_{it} = m_{it} \exp(\eta_{it})$. The linear predictor η_{it} will be decomposed additively into components depending on space, time, or both. For the spatial component we will adopt the standard Besag *et al.* (1991) model with a spatially unstructured and structured component, $\boldsymbol{\nu}$ and $\boldsymbol{\psi}$, say. The unstructured random effects ν_i, $i = 1, \ldots, I$, are assumed to be independent mean-zero normally distributed with unknown variance σ_ν^2. To account for the fact that geographically close regions often have similar incidence rates the spatially structured component $\boldsymbol{\psi}$ is modeled as an intrinsic Gaussian Markov

random field (IGMRF) (Rue and Held, 2005). Hence, the joint prior density of $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_I)^T$ can be written as

$$\pi(\boldsymbol{\psi}|\sigma^2_{\boldsymbol{\psi}}) \propto \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\psi}}}\sum_{i\sim i'}(\psi_i - \psi_{i'})^2\right), \tag{4}$$

where the sum in Equation (4) includes all pairs of adjacent regions i and i'. Introducing a so-called structure matrix $\boldsymbol{R}_{\boldsymbol{\psi}}$, the prior density (4) can be written compactly as

$$\pi(\boldsymbol{\psi}|\sigma^2_{\boldsymbol{\psi}}) \propto \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\psi}}}\boldsymbol{\psi}^T\boldsymbol{R}_{\boldsymbol{\psi}}\boldsymbol{\psi}\right).$$

This is the general form of an (I)GMRF and all other (I)GMRFs in this paper can be written in the same way. For example, the prior on $\boldsymbol{\nu}$ corresponds to a GMRF with structure matrix being simply the identity matrix $\boldsymbol{I}$. See Section 3.2 for further details on the form of $\boldsymbol{R}$ for temporal and spatio-temporal components. For the specific IGMRF (4) the diagonal entries of $\boldsymbol{R}_{\boldsymbol{\psi}}$ are equal to n_i, the number of neighbors of region i. The off-diagonal elements are equal to -1 if $i \sim i'$, and zero otherwise.

To model data in space and time Bernardinelli *et al.* (1995b) proposed a Bayesian model with parametric time trends. A main linear time trend and a so-called differential time trend for each region i are added to the spatially structured and unstructured components. The linear predictor can thus be written as

$$\eta_{it} = \mu + \nu_i + \psi_i + (\phi + \varphi_i)\cdot t. \tag{5}$$

The parameter ϕ represents an overall linear time trend whereas $\boldsymbol{\varphi}$ captures the interaction between the linear time trend and the regional effect $\boldsymbol{\nu}$ and $\boldsymbol{\psi}$, respectively. The formulation has similarities to the so-called random slope model, popular in the analysis of longitudinal data (Carvalho and Knorr-Held, 2003). The effect φ_i is called differential trend of the ith region, since it can be interpreted as the amount by which the time trend of region i differs from the overall time trend ϕ. For example, a negative φ_i indicates a region with a slope that is less steep than the overall time trend ϕ. As a result, $\mu + \nu_i + \psi_i$ can be interpreted as the spatial intercept of region i while $\phi + \varphi_i$ represents its slope.

According to Bernardinelli *et al.* (1995b) the prior specification of $\boldsymbol{\varphi}$ can be identical to that used for $\boldsymbol{\nu}$ and $\boldsymbol{\psi}$, i.e. either spatially structured or unstructured. To avoid high correlation between ν_i, ψ_i and φ_i, the time variable t should be centered around zero. Bernardinelli *et al.* (1995b) suggest to explicitly allow for correlation between ν_i (or ψ_i) and φ_i using a bivariate normal distribution. However, in the application described in Section 4 the correlation was found to be negligible after time centering. Hence, only models without explicit incorporation of correlation have been used in this application. In Model 1 (M1) the differential trend $\boldsymbol{\varphi}$ in Equation (5) is specified as an i.i.d. mean-zero normally distributed effect with structure matrix $\boldsymbol{R}_{\boldsymbol{\varphi}}$ equal to the identity matrix $\boldsymbol{I}$ whereas in Model 2 (M2) the prior of $\boldsymbol{\varphi}$ is specified as in Equation (4).

3.2. Models with nonparametric time trend

A limitation of the models described above is the assumption of a linear time trend in each region. A natural extension is to drop linearity and to assume a nonparametric model. We follow the framework outlined in Knorr-Held (2000) to specify prior distributions for the main time trend and the spatio-temporal interaction terms.

According to Knorr-Held (2000) the linear predictor of a nonparametric, additive space–time model can be written as

$$\eta_{it} = \mu + \nu_i + \psi_i + \gamma_t + \beta_t + \delta_{it}. \tag{6}$$

The spatial terms $\boldsymbol{\nu}$ and $\boldsymbol{\psi}$ are defined as in Section 3.1. The parameter $\boldsymbol{\gamma}$ represents an unstructured temporal effect which assumes no temporal structure a priori. An independent mean-zero normal prior with unknown variance σ^2_{γ} is used for $\boldsymbol{\gamma}$. In contrast, $\boldsymbol{\beta}$ displays temporal structure and follows a random walk of first order (RW1). Its prior density can be written as

$$\pi(\boldsymbol{\beta}|\sigma^2_{\beta}) \propto \exp\left(-\frac{1}{2\sigma^2_{\beta}}\sum_{t=2}^{T}(\beta_t - \beta_{t-1})^2\right). \tag{7}$$

The form of the respective structure matrix $\boldsymbol{R}_{\beta}$ is given in Knorr-Held (2000). The terms $\boldsymbol{\nu}$, $\boldsymbol{\psi}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$ are called main effects in the following while $\boldsymbol{\delta}$ represents space–time interactions. The specification of the prior on $\boldsymbol{\delta}$ depends on the spatial and temporal main effects which are assumed to interact. Four types of interactions are proposed in Knorr-Held (2000), a summary can be found in Table 2. Each interaction type can be interpreted in a different way. As an example, a Type II interaction is suitable if the temporal trends are different from region to region, but do not have any structure in space. If the data are given on a fine spatial grid it might be suitable to assume that the interaction effect is also structured in space. This means that not only the temporal and spatial neighbors enter in the full conditional prior distribution of a particular component of $\boldsymbol{\delta}$, but also the temporal neighbors of the spatial neighbors. This can be incorporated in the model with a Type IV interaction prior.

Table 2. Model acronym, specification and rank deficiency for four possible types of space–time interaction as proposed in Knorr-Held (2000)

Model (RW1/RW2)	Space–time Interaction	$\boldsymbol{R}_\delta$	Rank of $\boldsymbol{R}_\delta$	
			RW1 for $\boldsymbol{\beta}$	RW2 for $\boldsymbol{\beta}$
M3/7	Type I	$\boldsymbol{R}_\nu \otimes \boldsymbol{R}_\gamma$	$I \cdot T$	$I \cdot T$
M4/8	Type II	$\boldsymbol{R}_\nu \otimes \boldsymbol{R}_\beta$	$I \cdot (T-1)$	$I \cdot (T-2)$
M5/9	Type III	$\boldsymbol{R}_\psi \otimes \boldsymbol{R}_\gamma$	$(I-1) \cdot T$	$(I-1) \cdot T$
M6/10	Type IV	$\boldsymbol{R}_\psi \otimes \boldsymbol{R}_\beta$	$(I-1) \cdot (T-1)$	$(I-1) \cdot (T-2)$

As proposed in Clayton (1996) and Knorr-Held (2000) the structure matrix $\boldsymbol{R}_\delta$ for the prior of $\boldsymbol{\delta}$ can be obtained as the Kronecker product of the interacting main effects, e.g. $\boldsymbol{R}_\delta = \boldsymbol{R}_\nu \otimes \boldsymbol{R}_\gamma$ for a Type I interaction. For each interaction type the respective form of the joint density of $\boldsymbol{\delta}$ for $\boldsymbol{\beta}$ specified as an RW1 can be found in Knorr-Held (2000). The rank of the structure matrix obtained by the Kronecker product depends on the kind of interacting effects, see Table 2. Similar time-space interaction models have been used by several authors and in different applications, e.g. for age–period–cohort models (Lagazio *et al.*, 2003) and the joint analysis of two or more diseases (Richardson *et al.*, 2006).

Instead of using an RW1 prior for the temporal trend $\boldsymbol{\beta}$ a random walk of second order (RW2) can also be used. Especially if the data have a pronounced linear trend this might be appropriate. This modification has also been applied to a dataset on cancer mortality in West Germany in Schmid and Held (2004). The density of a random walk of second order is defined as

$$\pi(\boldsymbol{\beta}|\sigma_\beta^2) \propto \exp\left(-\frac{1}{2\sigma_\beta^2}\sum_{t=3}^{T}(\beta_t - 2\beta_{t-1} + \beta_{t-2})^2\right),$$

see Rue and Held (2005) for further details. The respective structure matrix $\boldsymbol{R}_\beta$ of this prior has the form

$$\boldsymbol{R}_\beta = \begin{bmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{bmatrix}.$$

For an interaction of Type I and III the main effect $\boldsymbol{\beta}$ now follows an RW2, but the specification of the interaction term $\boldsymbol{\delta}$ does not change. For an interaction of Type II and IV the prior for $\boldsymbol{\delta}$ can be written as

$$\pi(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{t=3}^{T}\sum_{i=1}^{I}(\delta_{it} - 2\delta_{i,t-1} + \delta_{i,t-2})^2\right)$$

and

$$\pi(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{t=3}^{T}\sum_{i\sim i'}((\delta_{it} - 2\delta_{i,t-1} + \delta_{i,t-2}) - (\delta_{i',t-2} - 2\delta_{i',t-1} + \delta_{i't}))^2\right),$$

respectively. The rank deficiency of the corresponding structure matrices $\boldsymbol{R}_\delta$ is higher than for models with an RW1 trend, see Table 2.

3.3. Specification and incorporation of linear constraints

In the nonparametric setting the parameters of the resulting models are not identifiable. Already the main effects $\boldsymbol{\psi}$ and $\boldsymbol{\beta}$ have to be centered around zero to ensure identifiability of the intercept μ. This is done by setting a sum-to-zero constraint on all ψ_i's and β_i's.

To ensure identifiability of the interaction term $\boldsymbol{\delta}$, specific sum-to-zero constraints have to be used. Only for a Type I interaction no additional constraints are necessary as this prior does not induce a rank deficiency, compare Table 2. If these constraints are not

incorporated then the interaction terms are confounded with the main time effect $\boldsymbol{\beta}$. The respective constraints can be derived using the results of Rue and Held (2005,Chapter 3.2). The details are as follows:

The vector $\boldsymbol{\delta}$ follows a so-called intrinsic Gaussian Markov random field (IGMRF). An IGMRF is improper, i.e. its precision matrix or, equivalently, its structure matrix $\boldsymbol{R}$ is not of full rank. Its improper density is denoted by $\pi^*(\boldsymbol{\delta})$ in the following. The order of an IGMRF is defined as the rank deficiency of its structure matrix. For the space–time interaction term the rank deficiency can be derived from Table 2. According to Rue and Held (2005) the vector $\boldsymbol{\delta}$ can always be decomposed into two parts, with the first part lying in the null space $\boldsymbol{\delta}^{\|}$ spanned by the structure matrix $\boldsymbol{R}$ and the second part $\boldsymbol{\delta}^{\perp}$ being orthogonal to this null space. It can be shown that the prior of $\boldsymbol{\delta}$ is invariant to the addition of any vector belonging to the null space of $\boldsymbol{R}$. Therefore, the improper density $\pi^*(\boldsymbol{\delta})$ can also be written as

$$\pi^*(\boldsymbol{\delta}) = \pi(\boldsymbol{\delta}|\boldsymbol{A}\boldsymbol{\delta} = \boldsymbol{e})$$

with $\boldsymbol{A}\boldsymbol{\delta} = \boldsymbol{e}$ denoting linear constraints on $\boldsymbol{\delta}$ where $\boldsymbol{A}$ is given by those eigenvectors of $\boldsymbol{R}$ which span the null space. Hence, the identifiability of $\boldsymbol{\delta}$ can be ensured by computing the null space of the respective structure matrix $\boldsymbol{R}$ and using the obtained eigenvectors as linear constraints for the estimation of $\boldsymbol{\delta}$. As a consequence the number of linear constraints which are necessary is always equal to the rank deficiency of $\boldsymbol{R}$ (see Table 2). This is a general result and can be applied in an automatic fashion to any type of interaction term to avoid confounding with the main effects. It should be noted that the number of required linear constraints depends on the total number of regions I and timepoints T and the chosen interaction type and may become quite large. Details on how to set these constraints using `INLA` can be found in Section 3.4.

Now consider the case in which $\boldsymbol{\beta}$ is specified as an RW1. In Knorr-Held (2000) specific sum-to-zero constraints have been proposed to ensure the identifiability of interaction terms of Type II, III, and IV. Using elementary matrix calculations it can be shown that these constraints are equivalent to those defined using the eigenvectors of the null space. Schmid and Held (2004) used the same sum-to-zero constraints for the case of an RW2 model for $\boldsymbol{\beta}$. However, they do not correspond to those based on the eigenvectors of $\boldsymbol{R}$ which span the null space.

3.4. Details on the implementation using INLA

In the following section some details on the implementation of spatio-temporal models using `INLA` are discussed. A comprehensive summary of all features can be found in the INLA manual (Martino and Rue, 2009) and the `R` help of the `INLA` package.

As noted in Section 3.3 interactions of Type II, III, and IV require the specification of a user-defined structure matrix $\boldsymbol{R}_{\boldsymbol{\delta}}$. Furthermore, linear constraints have to be set. As an example consider Model 4 (see Table 2) which incorporates an interaction of Type II. The time trend $\boldsymbol{\beta}$ is modeled as an RW1. The structure matrix of $\boldsymbol{\delta}$ is blockdiagonal with blocks specified analog to an RW1. To pass this matrix to `INLA` a list `Cmatrix <- list(i=c(),j=c(),Cij=c())` containing the non-zero entries of the structure matrix has to be defined as

$$\begin{array}{ccc} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & 2 \\ 2 & 3 & -1 \\ \vdots & \vdots & \vdots \end{array}$$

The first column contains the row index i, the second column the column index j and the third column the actual entry of the (i, j)th element of the structure matrix. All indexes must start with 1. Additionally, the effect must be specified as `model="generic0"`. The constraints on $\boldsymbol{\delta}$ can be passed to `INLA` using the `extraconstr`-option. All k additional constraints have to be arranged row-wise in a $k \times n$-matrix $\boldsymbol{A}$ in such a way that they fulfill the equation

$$\boldsymbol{A}\boldsymbol{\delta} = \boldsymbol{e}.$$

In `R` $\boldsymbol{A}$ and $\boldsymbol{e}$ have to be assigned to a list `clist <- list(A=A,e=e)`.

As noted in Section 2, the inclusion of linear constraints on the latent Gaussian field slows down the computation, if the simplified or the full Laplace approximation is used. Since the number of such constraints is large for the model introduced in Section 3.2, the Gaussian approximation was used for parameter estimation in Section 4. Note that this may result in a loss of accuracy of the estimated posterior distributions.

3.5. Priors

In a Bayesian framework prior distributions for all unknown variance parameters have to be specified. Here, inverse gamma distributions were used as priors. The priors for σ_{ψ}^2 and σ_{ν}^2 were adjusted for the fact, that σ_{ψ}^2 denotes a conditional variance whereas σ_{ν}^2 is interpretable as marginal variability, see Appendix A. Following Bernardinelli *et al.* (1995a) priors were chosen as inverse Ga(1, 0.018) and Ga(1, 0.01) distribution, respectively. In the parametric setting the prior for σ_{δ}^2 was chosen analog to the interacting spatial effect. In the nonparametric setting an inverse Ga(1, 0.01) was chosen for σ_{β}^2 (RW1), σ_{γ}^2 and σ_{δ}^2. If $\boldsymbol{\beta}$ is modeled as a RW2 the prior for σ_{β}^2 has to be chosen with care as the results can be sensitive to that choice. See Natario and Knorr-Held (2003) and Schrödle and Held (2009) for

Table 3. DIC for all models including a parametric time trend

	Linear time trend		
Model	$\bar{D}$	p_D	DIC
M1	3214	154	3368
M2	3214	144	3358

Table 4. DIC for all models including a nonparametric time trend

		RW1			RW2		
Model	Space–time interaction	$\bar{D}$	p_D	DIC	$\bar{D}$	p_D	DIC
M3/7	Type I	2834	457	3291	2837	452	3288
M4/8	Type II	2900	318	3218	3061	194	3254
M5/9	Type III	2884	395	3279	2885	397	3282
M6/10	Type IV	2922	288	3211	3069	191	3261

a discussion of this issue. In this application an inverse Ga(1, 0.00005) prior was chosen for σ_β^2 and σ_δ^2, respectively. Finally, for the fixed effects μ and β Gaussian priors with mean zero and variance 1000 are used.

4. RESULTS FOR SALMONELLOSIS IN CATTLE, 1991–2008

The following sections discuss the results obtained through applying the described models to Salmonellosis counts in cattle in Switzerland, 1991–2008.

4.1. Model choice

DIC values for the parametric models from Section 3.1 are given in Table 3. The fit of the two models is identical, but the number of effective parameters, i.e. the model complexity, is smaller for Model 2. Hence, the model with a spatially structured differential trend $\boldsymbol{\varphi}$ is preferred. The result for the nonparametric models including a RW1 time trend is similar (see Table 4): The model with a Type IV space–time interaction has the smallest DIC value. Model 4 and 6 assuming a space–time interaction with temporal structure (Type II/IV) have a worse fit than Models 3 and 5, but the estimated model complexity is lower. All models including an RW2 time trend have a higher DIC value than the respective model including an RW1, except for Model 7. Interestingly, Model 8 and 10 have a lower p_D than Models 4 and 6. This is due to the stronger dependency structure incorporated in the interaction terms when using an RW2. In general, an RW1 specification for $\boldsymbol{\beta}$ seems more appropriate than an RW2. Comparing the models with parametric and nonparametric time trend the nonparametric models are clearly preferred. Therefore, Model 6 is chosen as the best model.

4.2. Results

The estimated relative risk of Salmonellosis is shown in Figure 1(a) for the best model (M6). The differences between the southern and northern parts of Switzerland are quite large. In some regions the incidence of Salmonellosis is even 11 times as large as on average. Reasons for this strong spatial pattern are unclear. It is striking that the incidence of Salmonellosis is low especially in the mountainous areas in the southern part of Switzerland.

The linear time trend is estimated as -0.04 by the best parametric model (M2). The respective 95%-credible interval is $(-0.07; -0.02)$. This indicates a clearly decrease in number of reported cases from 1991 to 2008. The sum of the overall linear trend ϕ and the differential trend φ_i for each region is shown in Figure 1(b). Regions with a time trend that is significantly different from the main time trend are marked with white borders. Especially in central Switzerland there are regions where the incidence has been decaying even stronger than in the rest of the country. The number of reported cases has been rising in only 22 out of 185 regions, which are mostly located in the North-West. In Figure 2 the estimated nonparametric time trends for Models 6 and 10 are shown additionally to the linear time trend (M2). Models 6 and 10 assume a Type IV space–time interaction and $\boldsymbol{\beta}$ is modeled as an RW1 and RW2 term, respectively. The left plot shows that there is a strong nonparametric pattern in the time trend. The result for the RW2 model is tilted towards the linear trend as it smoothes towards a line rather than a constant term. For the nonparametric models the time trend can be split into its i.i.d. and RW1/RW2 component. The RW1 component of Model 6 is clearly nonparametric while the RW2 part of Model 10 is fairly smooth. In contrast, the i.i.d. trend exhibits more deviations from 0 for the RW2 model, especially for the years 1991–1997.

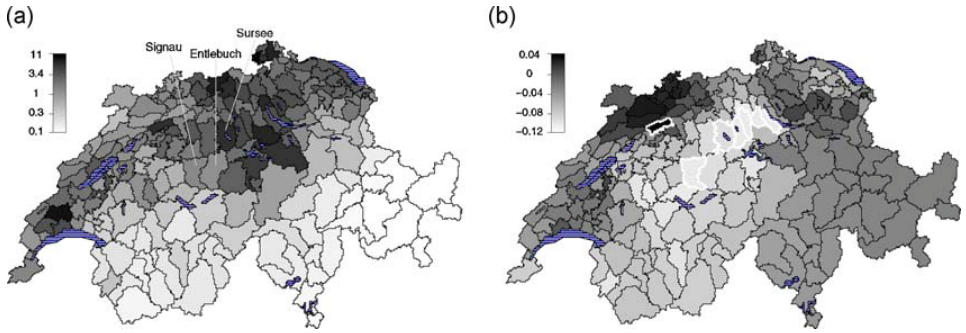


Figure 1. (a) Relative risk of Salmonellosis (M6); regions for which the linear predictor is plotted in Figure 3 are marked by arrows (b) Estimated time trend $\phi + \varphi_i$ on log scale for each region for the parametric model (M2); the estimated overall time trend is equal to -0.04 and marks the center of the color scale. Only 22 out of 185 region exhibit an increasing number of reported cases

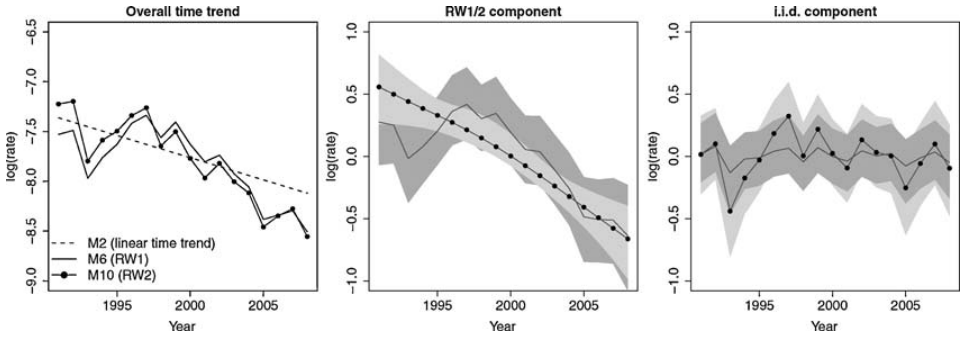


Figure 2. Overall time trend (left plot) for M2 (linear time trend, dashed), M6 (including RW1, solid) and M10 (including RW2, solid with dots). For the nonparametric models (M6/M10) the time trend can be split in its two components (RW1/2 and i.i.d component); they are shown in the center and right plot including 95%- credible intervals

To illustrate the implications of a special type of space–time interaction on the results the linear predictor on log scale is shown in Figure 3 for three Swiss regions. The regions are located in central Switzerland, see Figure 1(a). Signau is adjacent to Entlebuch in the West and Sursee is adjacent to Entlebuch in the East. The upper panel shows the results for Model 4 and 6 which both include a RW1 interaction term but assume a Type II and IV space–time interaction, respectively. Additionally, the log observed counts are shown. As there are no reported cases in many regions and years this value is $-\infty$ for many observations. In the Type IV setting the log incidence in Entlebuch for the years 2001–2008 is decaying stronger than for the Type II model. This is due to the fact that Signau and Sursee exhibit a strong decreasing time trend (see Figure 1(b)) which is inherited to Entlebuch in the Type IV interaction setting. During 1994 and 1998 there is a pronounced bump in Salmonellosis incidence in the region Sursee. Again, this is clearly inherited to the Entlebuch in the Type IV setting. The lower panel shows the respective results for the RW2 models. Here, the Type II and IV interaction model are more similar and credible intervals are more narrow in general. Furthermore, the estimated trends are smoother than for the RW1 models.

5. DISCUSSION

The goal of this paper was to show that a wide range of spatio-temporal disease mapping models can easily be fit using INLA. All models were applied to counts on Salmonellosis in cattle collected in Switzerland during the years 1991–2008. Using DIC the best model was chosen and conclusions concerning the pattern of Salmonellosis in Switzerland were drawn.

An advantage of INLA is its easy usability. When the R interface is used all models can be specific in a modular way similar to, e.g., generalized linear models. Additionally, all results can be processed in R.

Using INLA it is also straightforward to incorporate fixed covariates in the spatio-temporal disease mapping model to explain the heterogeneity in disease risk (Clayton *et al.*, 1993). See Schrödle and Held (2009) for an application in a purely spatial context.

In the nonparametric setting of spatio-temporal models originally proposed by Knorr-Held (2000) it is important to set linear constraints with respect to the space–time interaction term to ensure identifiability of the parameters and avoid confounding with the main effects. These constraints are equivalent to the eigenvectors which span the null space of the structure matrix of the interaction term and can be computed in an automatic fashion. Depending on the total number of regions and timepoints and the chosen interaction type the number of the required constraints can get quite high.

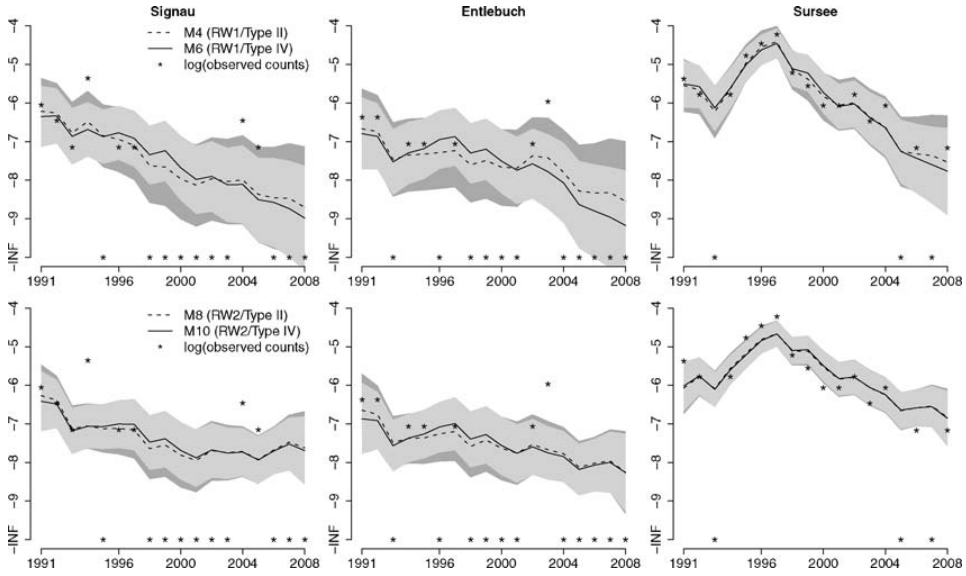


Figure 3. Plots of the linear predictor on logarithmic scale for three neighboring regions; for their location see Figure 1(a). The upper panel shows the results for M4 (including RW1/Type II)(dashed) and M6 (including RW1/Type IV)(solid) and the actually observed counts on log scale (*). The lower panel shows the according results for models including a RW2 time trend (M8/M10). All estimates are shown including 95%- credible intervals

Acknowledgment

Financial support by the Swiss Federal Veterinary Office (BVET) is gratefully acknowledged.

REFERENCES

Assunção R, Reis I, Oliveira C. 2001. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine* **20**(15): 2319–2335.
Banerjee S, Carlin B, Gelfand A. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC.
Bernardinelli L, Clayton D, Montomoli C. 1995a. Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* **14**: 2411–2431.
Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M. 1995b. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* **14**: 2433–2443.
Besag J, York J, Mollie A. 1991. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**(1): 1–59.
Carvalho M, Knorr-Held L. 2003. Modelling discrete time survival data with random slopes: evaluating haemodialysis centres. *Statistics in Medicine* 3543–3555.
Clayton D. 1996. Generalized linear mixed models. *In Markov Chain Monte Carlo in Practice* Gilks W, *et al*. (eds). Chapman & Hall: London; 275–301.
Clayton D, Bernardinelli L, Montomoli C. 1993. Spatial correlation in ecological analysis. *International Journal of Epidemiology* **22**(6): 1193–1202.
Czado C, Gneiting T, Held L. 2009. Predictive model assessment for count data. *Biometrics* **65**(4): 1254–1261.
Gneiting T, Raftery AE. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**(477): 359–378.
Held L. 2004. Simultaneous posterior probability statements from Monte Carlo output. *Journal of Computational and Graphical Statistics* **13**(1): 20–35.
Held L, Schrödle B, Rue H. 2010. Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures—Festschrift in Honour of Ludwig Fahrmeir*, Kneib T, Tutz G (eds). Physica-Verlag: Heidelberg. 91–110
Knorr-Held L. 2000. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**: 2555–2567.
Knorr-Held L, Rue H. 2002. On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics* **29**(4): 597–614.
Lagazio C, Biggeri A, Dreassi E. 2003. Age-period-cohort models and disease mapping. *Environmetrics* **14**: 475–490.
Martínez-Beneito M, López-Quilez A, Botella-Rocamora P. 2008. An autoregressive approach to spatio-temporal disease mapping, *Statistics in Medicine* **27**: 2874–2889.
Martino S, Akerkar R, Rue H. 2010. Approximate Bayesian inference for survival models. *Technical report*, Norwegian University of Science and Technology: Trondheim.
Martino S, Rue H. 2009. Implementing approximate Bayesian inference using integrated nested Laplace approximation: a manual for the INLA program. Available from: http://www.math.ntnu.no/∼hrue/GMRFLib
Natario I, Knorr-Held L. 2003. Non-parametric ecological regression and spatial variation. *Biometrical Journal* **45**(6): 670–688.
Paul M, Riebler A, Bachmann LM, Rue H, Held L. 2010. Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine*. **29**:1325–1339.
R Development Core Team. 2005. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: http://www.R-project.org.
Richardson S, Abellan J, Best N. 2006. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire. *Statistical Methods in Medical Research* **15**(4): 385–407.
Rue H. 2005. Marginal variances for Gaussian Markov random fields. Available from: http://www.math.ntnu.no/preprint/statistics/2005/S1-2005.pdf.

Rue H, Held L. 2005. *Gaussian Markov Random Fields*. Chapman & Hall/CRC.
Rue H, Martino S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference* **137**: 3177–3192.
Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B* **71**: 319–392.
Schmid V, Held L. 2004. Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* **60**: 1034–1042.
Schrödle B, Held L. 2010. A primer on disease mapping and ecological regression using INLA, *Computational Statistics*. DOI: 10.1007/s00180-010-0208-2.
Schrödle B, Held L, Riebler A, Danuser J. 2010. Using INLA for the evaluation of veterinary surveillance data from Switzerland: a case study. *Journal of the Royal Statistical Society*, Series C. Accepted.
Spiegelhalter D, Best N, Carlin B, van der Linde A. 2002. Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Series B* **64**(4): 583–639.
Tierney L, Kadane JB. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**(393): 82–86.
Waller L, Carlin B, Xia H, Gelfand A. 1997. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**: 607–617.

Appendix A: CHOOSING THE PRIORS FOR σ_v^2 AND σ_ψ^2

As noted in Section 3.5, the priors for σ_v^2 and σ_ψ^2 have to be specified with care. Usually inverse gamma priors with default parameters ξ_1 and ξ_2 are adopted. If the data are sparse as in this application the priors should be chosen carefully.

The variance σ_v^2 of the spatially unstructured effect is interpretable as marginal variability whereas σ_ψ^2 controls the variability of the random effect conditional upon the random effects in neighboring regions (Bernardinelli *et al.*, 1995a). To adopt the same amount of smoothing for both components, the relation between σ_ψ^2 and σ_v^2 has to be explored. For each region this association depends on the number n_i of adjacent regions, but also on the actual graph of the considered area.

Firstly, the marginal variances of the IGMRF $\boldsymbol{\psi}$ have to be calculated. According to Rue and Held (2005) $\boldsymbol{\psi}$ has an improper density $\pi^*(\boldsymbol{\psi}|\sigma_\psi^2)$ of the form (4) with a singular precision matrix $\boldsymbol{Q}$ with entries

$$Q_{ii'} = \frac{1}{\sigma_\psi^2} \cdot \begin{cases} n_i & \text{, if } i = i' \\ -1 & \text{, if } i \sim i' \\ 0 & \text{, otherwise.} \end{cases} \tag{8}$$

Using the results already derived in Section 3.3 the density of this IGMRF can also be written as the density of a proper GMRF under a linear constraint

$$\pi(\boldsymbol{\psi}|\boldsymbol{A}\boldsymbol{\psi} = \boldsymbol{e}).$$

For $\boldsymbol{\psi}$ the components of the linear constraint are $\boldsymbol{A} = (1, \ldots, 1)^T$ and $\boldsymbol{e} = \boldsymbol{0}$, i.e. all components ψ_i sum up to zero. Using these results the marginal variances of $\boldsymbol{\psi}$ can be calculated as described in Rue (2005).

To obtain the average ratio between the conditional and marginal standard deviation of $\boldsymbol{\psi}$ which is needed to adjust the priors (Bernardinelli *et al.*, 1995a) the parameter σ_ψ^2 can be fixed at 1, since it only involves a change of scale (Rue and Held, 2005). The ratio was found to be 0.75 for the graph from Switzerland. This finding was applied to a prior distribution (Prior B) suggested in Bernardinelli *et al.* (1995a). As a result inverse Gamma distributions Ga(1, 0.01) and Ga(1, 0.018) were chosen as priors for σ_v^2 and σ_ψ^2.

PAPER V

Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: A case-study

Birgit Schrödle, Leonhard Held, Andrea Riebler & Jürg Danuser

Paper published in *Journal of the Royal Statistical Society - Series C*, 2011, **60** (2), 261-279.

Journal of the Royal Statistical Society
SERIES C Applied Statistics
Appl. Statist. (2011)
60, *Part* 2, *pp.* 261–279

Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study

Birgit Schrödle, Leonhard Held and Andrea Riebler

University of Zurich, Switzerland

and Jürg Danuser

Federal Veterinary Office, Berne, Switzerland

[Received June 2009. Final revision August 2010]

Summary. Spatiotemporal disease mapping models have been used extensively to describe the pattern of surveillance data. They are usually formulated in a hierarchical Bayesian framework and posterior marginals are not available in closed form. Hence, the standard method for parameter estimation is Markov chain Monte Carlo algorithms. A new method for approximate Bayesian inference in latent Gaussian models using integrated nested Laplace approximations has recently been proposed as an alternative. This approach promises very precise results in short computational time. The aim of the paper is to show how integrated nested Laplace approximations can be used as an inferential tool for a variety of spatiotemporal models for the analysis of reported cases of bovine viral diarrhoea in cattle from Switzerland. Conclusions concerning the problem of under-reporting in the data are drawn via a multilevel modelling strategy. Furthermore, a comparison with Markov chain Monte Carlo methods with regard to the accuracy of the parameter estimates and the usability of both approaches in practice is conducted. Approaches to model choice using integrated nested Laplace approximations are also presented.

Keywords: Disease mapping; Integrated nested Laplace approximations; Leave-one-out cross-validation; Spatiotemporal models

1. Introduction

Spatiotemporal disease mapping models have been used extensively to describe the spatial and temporal pattern of registry data. Various specifications of the spatial and temporal trends and the space–time interaction term have been proposed in the literature (Bernardinelli *et al.*, 1995b; Knorr-Held, 2000; Lagazio *et al.*, 2003). From an inferential point of view, this class of models is formulated within a hierarchical Bayesian framework (Besag *et al.*, 1991; Banerjee *et al.*, 2004). As, in general, posterior marginals are not available in closed form, Markov chain Monte Carlo (MCMC) algorithms have been used for parameter estimation so far. But the often complex dependence structure in spatiotemporal models requires specific algorithms to obtain reliable estimates (Knorr-Held and Rue, 2002; Schmid and Held, 2004). Furthermore, MCMC methods may lead to a large Monte Carlo error and the computation time can be long.

Address for correspondence: Birgit Schrödle, Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, Zurich, Switzerland.
E-mail: birgit.schroedle@ifspm.uzh.ch

 0035–9254/11/60261

Recently, an approximate method for parameter estimation in specific Bayesian hierarchical models, so-called latent Gaussian models, has been proposed in Rue *et al.* (2009). This method uses integrated nested Laplace approximations (INLAs) to approximate the posterior marginals of interest. Since spatiotemporal disease mapping models incorporate a latent Gaussian field, the INLA approach can be used for inference here. A major advantage of INLAs is that computational time is short and they can easily be used via the R library (R Development Core Team, 2005) `INLA` (Martino and Rue, 2009). Quantities for model criticism and comparison, e.g. the well-known deviance information criterion (DIC), are provided as standard. Furthermore, cross-validated diagnostic tools based on the predictive distribution can be obtained. These tools, namely the probability integral transform (PIT) and the logarithmic score (Gneiting and Raftery, 2007), have recently been applied to count data by Czado *et al.* (2009). A detailed comparison of these criteria with results from MCMC methods is given in Held *et al.* (2010).

In this paper, the scope of INLAs concerning spatiotemporal disease mapping is assessed by means of a case-study. Parameter estimates for a data set containing reported cases of bovine viral diarrhoea (BVD) in cows from Switzerland were obtained by using INLAs. Their accuracy is assessed via a comparison with results from MCMC techniques. Posterior samples were drawn by using auxiliary mixture sampling (Frühwirth-Schnatter *et al.*, 2009) and a second-order Taylor series expansion of the log-likelihood to obtain a suitable proposal for a Metropolis–Hastings algorithm (Rue and Held, 2005). Furthermore, the usability of INLAs and MCMC methods in terms of available software and computational time is discussed briefly.

BVD is a viral diarrhoea infection in cattle. It is one of the most widespread cattle diseases in Switzerland and causes damage of several million Swiss francs every year (Swiss Federal Veterinary Office, 2006). The reported cases, which were collected by the Swiss Federal Veterinary Office as part of routine surveillance from 2003 to 2007, show an increasing trend (Table 1). However, it is well documented that case reporting data can potentially be biased owing to limited case detection or low reporting motivation (Doherr and Audige, 2001). This suspicion is confirmed by the fact that the number of reported BVD cases varies heavily throughout the country. Nevertheless, there is no obvious reason for such large variability in these data as BVD is a slowly spreading viral disease. Several of the 185 Swiss regions reported no cases of BVD during the time period whereas some regions had a stable, high number of reported cases. Additionally, the strong rise in reported cases gives reason to doubt that the temporal heterogeneity is due to a real increase in prevalence of disease.

Switzerland is a confederation of 26 cantons. Each canton consists of one or more regions. The cantonal veterinary authorities are responsible for the realization of federal veterinary legislation. Hence, cantons build a second, coarser spatial grid. It is suspected that the system of case registration is highly influenced by the affiliation of a region to a certain canton. This heterogeneity could be caused by a cantonal difference in incentives for a farmer to report a case, e.g. financial compensation in the case of a diseased animal, or a different practice in conveying disease information to farmers. Hence, multilevel models similar to those of Langford *et al.* (1998, 1999) addressing this issue are formulated and evaluated by using model choice.

Table 1. Number of reported BVD cases, 2003–2007

Year	2003	2004	2005	2006	2007
Number	141	172	239	406	712

This paper is organized as follows: Section 2 introduces variations of spatiotemporal models which are appropriate to investigate the spatial and temporal pattern in the data and to assess cantonal heterogeneity. In Section 3 the INLA approach is described; Section 4 discusses tools for model comparison that are returned by INLAs. All results with regard to case reporting for BVD are presented in Section 5 in detail. Various aspects of the comparison of INLAs and MCMC methods are discussed in Section 6. We close with some general results in Section 7.

2. Spatiotemporal models

In what follows we outline five models to describe the spatiotemporal pattern of the BVD data by using a disease mapping approach. As described in Section 1, Switzerland is a confederation of 26 cantons which consist of one or more regions: 184 in total. As the cantonal veterinary authorities are responsible for the implementation of federal veterinary legislation, it is of interest to investigate whether the cantonal affiliation of a region has a pronounced influence on case reporting. Hence, a cantonal effect is included in three of the models. This is done by using a multilevel approach such that variability in the response is attributed to different hierarchical levels (Langford *et al.*, 1998, 1999). Furthermore, models incorporating a linear as well as a non-parametric time trend are proposed. Space–time interactions that adjust for the fact that spatial units can behave differently from the main time trend when observed over a long time span are additionally included in each type of model. Models assuming a linear time trend have been proposed in Bernardinelli *et al.* (1995b) and developed further in Assunção *et al.* (2001). Non-parametric space–time interaction models have been introduced by Knorr-Held (2000) and used by several researchers and in different applications, e.g. for age–period–cohort models (Lagazio *et al.*, 2003; Schmid and Held, 2004) and the joint analysis of two or more diseases (Richardson *et al.*, 2006). The characteristics of all models that will be introduced in Sections 2.1 and 2.2 are summarized in Table 2.

2.1. Linear time trend

Since BVD cases are available for each Swiss region separately, this fine grid can be used as a basis for a spatial analysis. Additionally, data from the Principality of Liechtenstein are included (which counts as both a region and a canton). A case means that at least one diseased animal within one herd was detected. Under the rare disease assumption, it is assumed that the number of cases of disease y_{it} in region $i = 1, \ldots, I$ at time $t = 1, \ldots, T$ is Poisson distributed with parameter λ_{it}, which can be interpreted as the relative risk of the disease in region i at time t. Additionally, the number of herds m_{it} is included as an offset to adjust for the different number

Table 2. Characteristics of all models from Sections 2.1 and 2.2

Model	*Cantonal heterogeneity* α	*Time trend* β	*Space–time interaction* δ *(Knorr-Held, 2000)*
M1	✓	Linear	—
M2	✓	RW1	Type II
M3	✓	RW2	Type II
M4	—	RW1	Type IV
M5	—	RW2	Type IV

of herds at risk. Following the standard generalized linear mixed model formulation (Breslow and Clayton, 1993) with Poisson response, a logarithmic link is used.

To account for a linear time trend, Bernardinelli *et al.* (1995b) proposed a Bayesian spatiotemporal model. It can be seen as an extension of the standard model for disease mapping that was introduced by Besag *et al.* (1991). In the standard setting defined for one spatial level, a main linear time trend and a so-called differential trend for each area i are incorporated in the model as well as spatially structured and unstructured effects. In this application, where cantonal heterogeneity is considered as well, we assume that a cantonal effect α_j interacts with the linear time trend. Hence, the rate of cases over time can have a different slope for each canton. Reasons for this heterogeneity could be cantonal differences in incentives for a farmer to report a case, e.g. financial compensation in the case of a diseased animal, or a different practice in conveying disease information to farmers. The linear predictor of this model (M1) can be written as

$$\log(\lambda_{it}) = \eta_{it} = \log(m_i) + \mu + \nu_i + \alpha_{j(i)} + (\phi + \varphi_{j(i)})t \tag{1}$$

with $i = 1, \ldots, 185$, $t = 1, \ldots, 5$ and $j = 1, \ldots, 27$. The index $j(i)$ denotes the canton j which region i belongs to. The offset m_i was provided by the Swiss Federal Veterinary Office and is supposed to be constant in time. The vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_I)^{\mathrm{T}}$ is independent and identically mean 0 normally distributed with variance σ_ν^2. The ν_is account for differences between regions whereas the α_js model cantonal heterogeneity. In this model the similarity of the incidence between neighbouring regions is described via the cantonal term. Hence, it incorporates a two-level structure. The parameter ϕ represents the overall linear time trend. The term φ_j depicts the interaction between the linear time trend and the cantonal intercept α_j and is modelled as a random slope. Thus, $\phi + \varphi_j$ represents the individual time trend for canton j. Each φ_j can be interpreted as the amount by which the time trend of canton j differs from the overall trend ϕ. A prior distribution for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)^{\mathrm{T}}$ and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_J)^{\mathrm{T}}$ must be defined as well. Since it is assumed that the cantonal effects $\boldsymbol{\alpha}$ are independent for each canton, the differential trends $\boldsymbol{\varphi}$ are modelled in the same way (Bernardinelli *et al.*, 1995b). Furthermore, it is necessary to allow for correlation between the intercept and slope in a random-slope model (Hedeker and Gibbons (2006), section 4.4.2). A standard assumption is that $(\alpha_j, \varphi_j)^{\mathrm{T}}$ follows a bivariate normal distribution with mean 0 and some unknown precision matrix $\mathbf{P}$, to which a Wishart prior is assigned. Bernardinelli *et al.* (1995b) also proposed that the time variable t should be centred at 0 to avoid high correlation between the intercept and slope. We have followed this advice in our application. The specification of hyperpriors is discussed in Section 2.3.

2.2. Non-parametric time trend

In model M1, the time trend in log-incidence is taken as linear. This assumption can be relaxed by adopting a non-parametric setting as proposed in Knorr-Held (2000). Custom-made modifications of this general setting are formulated for the BVD data in what follows.

The second model M2 is the non-parametric analogue of model M1. In contrast with model M1 it includes a main time trend $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)^{\mathrm{T}}$ and an interaction $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{1T}, \delta_{21}, \ldots, \delta_{2T}, \ldots, \delta_{JT})^{\mathrm{T}}$ between canton and time to which specific prior distributions must be assigned. The linear predictor is

$$\eta_{it} = \log(m_i) + \mu + \nu_i + \alpha_{j(i)} + \beta_t + \delta_{j(i)t}. \tag{2}$$

Here, the α_js are modelled as independent and identically mean 0 normally distributed with variance σ_α^2. For $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ we use intrinsic Gaussian Markov random-field priors of the general form

$$\pi(\boldsymbol{\beta}|\sigma^2) \propto \sigma^{-\text{rank}(\mathbf{R})} \exp\Big(-\frac{1}{2\sigma^2}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{R}\boldsymbol{\beta}\Big), \tag{3}$$

including a so-called structure matrix $\mathbf{R}$ (Held and Rue, 2010). The main time trend is specified as a random walk (RW) of first order with structure matrix

$$\mathbf{R}_{\beta} = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & -1 & 2 & -1 & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 1 & \end{pmatrix}. \tag{4}$$

The assumption of temporal structure is plausible as the number of reported cases is constantly increasing over time. The joint prior density of $\boldsymbol{\beta}$ can be written as (Rue and Held, 2005)

$$\pi(\boldsymbol{\beta}|\sigma_{\beta}^2) \propto \exp\Big\{-\frac{1}{2\sigma_{\beta}^2}\sum_{t=2}^{T}(\beta_t - \beta_{t-1})^2\Big\}. \tag{5}$$

To specify the prior on $\boldsymbol{\delta}$ we consider the interacting spatial ($\boldsymbol{\alpha}$) and temporal ($\boldsymbol{\beta}$) main effects: since the cantonal effects $\boldsymbol{\alpha}$ are modelled as spatially unstructured, a so-called type II interaction prior (Knorr-Held, 2000) is used for $\boldsymbol{\delta}$, i.e. the interactions δ_{jt} in the different cantons follow independent RWs in time. Hence, the form of the resulting joint distribution for $\boldsymbol{\delta}$ is similar to expression (5), including an additional sum over all cantons:

$$\pi(\boldsymbol{\delta}|\sigma_{\delta}^2) \propto \exp\Big\{-\frac{1}{2\sigma_{\delta}^2}\sum_{j=1}^{J}\sum_{t=2}^{T}(\delta_{jt} - \delta_{jt-1})^2\Big\}. \tag{6}$$

Following Clayton (1996) and Knorr-Held (2000), its structure matrix can be obtained as the Kronecker product of the interacting main effects and has rank $J(T-1)$. To ensure identifiability of the main time trend $\boldsymbol{\beta}$, the δ_{jt}s must sum to 0 for each $j = 1, \ldots, J$.

Instead of a first-order RW prior for $\boldsymbol{\beta}$ an RW of second order can be used. This assumption might be appropriate for the BVD data which exhibit an increasing number of counts over the observed time period. A first-order RW trend smooths towards a constant whereas the second-order RW penalizes deviations from a linear trend. The structure matrices of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ and the linear constraints must be adapted appropriately; see Schmid and Held (2004) and Rue and Held (2005) for details. This new model, which includes a second-order RW main time trend and the respective interaction, is called model M3 in this application.

So far, all models proposed explicitly include cantonal heterogeneity. To investigate whether a cantonal component is necessary, models with regional effects only are considered as well. Similarities between neighbouring regions are now modelled by using an intrinsic Gaussian Markov random field for $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_I)^{\mathrm{T}}$ with prior density

$$\pi(\boldsymbol{\psi}|\sigma_{\psi}^2) \propto \exp\Big\{-\frac{1}{2\sigma_{\psi}^2}\sum_{i\sim i'}(\psi_i - \psi_{i'})^2\Big\}. \tag{7}$$

The sum in expression (7) includes all pairs of adjacent regions i and i'. The linear predictor of the resulting model M4 is given as

$$\eta_{it} = \log(m_i) + \mu + \psi_i + \nu_i + \beta_t + \delta_{it}. \tag{8}$$

In equation (8), the time trend $\boldsymbol{\beta}$ is modelled as a first-order RW. Since the Swiss regions build a fine spatial grid we assume (in contrast with the preceding models) that the interaction effects $\boldsymbol{\delta}$ are also spatially structured. This means that both the temporal and the spatial neighbours as well as the temporal neighbours of the spatial neighbours enter the conditional distribution of the Gaussian Markov random field. This assumption is appropriate if temporal trends are different from region to region but are more likely to be similar for adjacent regions. This can be incorporated in the model with a type IV interaction prior (Knorr-Held, 2000) of the form

$$\pi(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left\{-\frac{1}{2\sigma_\delta^2}\sum_{t=2}^{T}\sum_{i\sim i'}(\delta_{it}-\delta_{i't}-\delta_{i,t-1}+\delta_{i',t-1})^2\right\}. \quad (9)$$

The appropriate structure matrix can be obtained by the Kronecker product of the structure matrices (4) of the first-order RW term $\boldsymbol{\beta}$ and the structure matrix of the intrinsic Gaussian Markov random-field prior on $\boldsymbol{\psi}$. This model induces full dependence over time and space. The rank of the structure matrix is now $(I-1)(T-1)$. To avoid problems of identifiability, the δ_{it}s need to sum to 0 for each i and each t, i.e.

$$\sum_{i=1}^{I}\delta_{it}=0 \qquad \text{for all } t=1,\ldots,T,$$

$$\sum_{t=1}^{T}\delta_{it}=0 \qquad \text{for all } i=1,\ldots,I.$$

One of these $I+T$ constraints is redundant.

By analogy with the non-parametric models including cantonal heterogeneity, a fifth model M5 is fitted. In this model a second-order RW prior is assigned to $\boldsymbol{\beta}$ and the structure matrix of the interaction term $\boldsymbol{\delta}$ is obtained as the Kronecker product of the structure matrices of $\boldsymbol{\psi}$ and $\boldsymbol{\beta}$ (second-order RW).

In this application, a herd is the unit of analysis (see Section 2.1). Therefore, a large herd may be more likely to be a case than a small herd, as there are more animals at risk. In most Swiss regions the mean number of cows per herd is between 30 and 40. An ecological regression including the logarithm of the mean herd size as explanatory variable is conducted in Section 5.2 to investigate this issue.

2.3. Priors

Since the models are formulated in a Bayesian way, prior distributions must be assigned to all variance and precision components. In the parametric setting (1) a Wishart prior is assigned to the precision matrix $\mathbf{P}$ of the bivariate normal distribution for $(\alpha_j, \varphi_j)^{\mathrm{T}}$. The Wishart distribution $\mathrm{Wi}_2(l, \mathbf{L})$ has two components, namely the degrees of freedom l and the matrix $\mathbf{L}$. Here, they were chosen as $l=4$ and

$$\mathbf{L}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

a priori. For σ_ν^2 an inverse gamma prior IG(1,0.01) was used. The parameterization of the inverse gamma distribution is as in Natario and Knorr-Held (2003) and Rue *et al.* (2009).

In the non-parametric settings independent IG(1,0.01) priors were used for σ_ν^2, σ_α^2, σ_β^2 (first-order RW) and σ_δ^2 (as specified in expression (6)). In models M4 and M5 the prior of σ_ψ^2 was adjusted for the fact that it represents conditional variability on the same spatial level as σ_ν^2 and chosen as IG(1,0.018) (Bernardinelli *et al.*, 1995a). For the models including a second-order

RW specification of $\boldsymbol{\beta}$ (model M3 or M5), an IG(1,0.00005) prior was used for $\sigma^2_{\boldsymbol{\beta}}$ and $\sigma^2_{\boldsymbol{\delta}}$. For a discussion of the sensitivity with respect to this prior see Natario and Knorr-Held (2003).

3. Integrated nested Laplace approximations—a new approach for approximate Bayesian inference

INLAs are a recently proposed method for approximate Bayesian inference in structured additive regression models with latent Gaussian fields (Rue *et al.*, 2009).

Spatiotemporal models such as introduced in Section 2 fit into this framework and are built in a hierarchical fashion including three stages. The first stage is the observational model $\pi(\mathbf{y}|\mathbf{x})$. The second stage is the latent Gaussian field $\pi(\mathbf{x}|\boldsymbol{\theta})$ with precision matrix $\mathbf{Q}$, e.g. $\mathbf{x} = (\mu, \boldsymbol{\nu}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}$ for expression (2). It is typically controlled by a few hyperparameters $\boldsymbol{\theta}$ which are not necessarily Gaussian (third stage). All unknown variance parameters (e.g. σ^2_{α}) as specified in Section 2 enter $\boldsymbol{\theta}$.

For such models it is not possible to compute the posterior distributions analytically. Hence, MCMC methods have been used to obtain estimates so far, but they have some drawbacks: the computational time may be long if samples are highly correlated. Especially for models with a complex dependence structure within the Gaussian field such as proposed in Section 2, advanced MCMC algorithms are required to provide a reasonable sampler for the posterior marginals. This issue is discussed in more detail in Section 6. In contrast, INLAs provide accurate approximations to the posterior marginals in short computational time. In what follows we present the inference strategy briefly; for details refer to Rue *et al.* (2009).

The main goal is to estimate the marginal posterior distribution

$$\pi(x_i|\mathbf{y}) = \int_{\boldsymbol{\theta}} \underbrace{\pi(x_i|\boldsymbol{\theta}, \mathbf{y})}_{\text{part 2}} \underbrace{\pi(\boldsymbol{\theta}|\mathbf{y})}_{\text{part 1}} \, \mathrm{d}\boldsymbol{\theta} \tag{10}$$

given the data for each component x_i of the latent Gaussian field $\mathbf{x}$. Parts 1 and 2 of equation (10) are processed in an elaborate way. From $\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}) = \pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})\pi(\mathbf{y})$ it follows that part 1 of integral (10) can be approximated by

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left. \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{\mathrm{G}}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}, \tag{11}$$

which is the Laplace approximation of a marginal posterior distribution (Tierney and Kadane, 1986). In expression (11), $\tilde{\pi}_{\mathrm{G}}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ denotes the Gaussian approximation (Rue and Held, 2005) to $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\mathbf{x}$ for a given $\boldsymbol{\theta}$. To integrate out the uncertainty with respect to $\boldsymbol{\theta}$, it is essential to explore the properties of expression (11) and to find good evaluation points $\boldsymbol{\theta}_k$ for a numerical integration of equation (10). This is done by an iterative algorithm (Rue *et al.*, 2009). Additionally, an appropriate area weight Δ_k must be assigned to each $\boldsymbol{\theta}_k$ (see equation (12)).

For the approximation of part 2 in equation (10), three alternatives were proposed in Rue *et al.* (2009): a Laplace approximation, a simplified Laplace approximation and the simplest of these: the Gaussian approximation. Here, the distribution of a non-normal variable is approximated with a Gaussian distribution by matching the mode and the curvature at the mode (Rue and Held (2005), section 4.4.1). According to Rue and Martino (2007), this method often gives reasonable results but there can be errors in the location or due to the lack of skewness or both. Therefore, the approximations can be improved by applying the Laplace approximation also to $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$. This so-called full Laplace approximation is very precise. Rue *et al.* (2009) also

proposed an alternative method, the simplified Laplace approximation, which is based on a series expansion of the full Laplace approximation. This method takes less computation time than the full Laplace approximation and is equally accurate in many applications. Putting things together, we obtain

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k \tag{12}$$

as an approximation of the posterior marginal density (10).

As noted in Section 2.2, the incorporation of linear constraints on $\mathbf{x}$ is required in models M2–M5. This is possible by using INLAs but will slow down the computational time for the simplified and the full Laplace approximation, if the number of constraints is large.

From a user's point of view INLAs can be used in a modular way. The program `inla` which is written in C is bundled within an R library (R Development Core Team, 2005) called `INLA` which permits model specification and processing of the results directly in R. It can be downloaded freely from `http://www.r-inla.org` and is available for LINUX, Macintosh and Windows environments. All analyses within this paper were run by using the `INLA` package built on June 9th, 2010, on `inla` version 1.2.

A detailed comparison of INLA approximations and MCMC histograms for model M2 is given in Section 6.

4. Model comparison and calibration

An important feature of the INLA approach is that criteria for model choice and assessment of model calibration can be obtained directly from the `INLA` output (Rue *et al.* (2009), section 6.4). Even cross-validated quantities that are needed for computation of the logarithmic score and the PIT that are discussed below can be computed by INLAs without rerunning the model. Their accuracy in comparison with quantities that are obtained by MCMC methods is discussed in Held *et al.* (2010).

4.1. Deviance information criterion

The DIC is a popular criterion for Bayesian model selection. According to Spiegelhalter *et al.* (2002) it is the sum of the posterior mean of the deviance $\bar{D}$ and the number of effective parameters p_D. A low mean deviance indicates a good model fit, but it decreases with an increasing number of parameters. Hence, the effective number of parameters is added to penalize model complexity. So the model with the lowest DIC provides the best trade-off between model fit and complexity.

4.2. Logarithmic score

One approach for assessing the predictive performance of a model is to use cross-validated scoring rules which assign each model a numerical score based on the predictive distribution. Cross-validation means that one observation y_{it} is left out in each step of the validation process and the predictive distribution $P_{y_{it}} = \text{Prob}(Y_{it} \leqslant y_{it}|\mathbf{y}_{-it})$ based on the remaining observations is computed. For discrete Y_{it} the logarithmic score that is considered in Section 5.1 is defined as

$$\text{LS} = -\log(\pi_{y_{it}}), \tag{13}$$

where $\pi_{y_{it}} = \text{Prob}(Y_{it} = y_{it}|\mathbf{y}_{-it})$ denotes the cross-validated predictive probability mass at the observed count. Both $P_{y_{it}}$ and $\pi_{y_{it}}$ are available in `INLA`. According to Stone (1977), the cross-validated mean logarithmic score is asymptotically equivalent to the Akaike information

criterion if the observations are independent. Here, scoring rules are negatively oriented, which means that, the smaller the score, the better the predictive power of the model. An attractive feature of this measure is that it can be applied to parametric and non-parametric settings and does not require models to be nested, nor to be related in any way (Gneiting and Raftery, 2007).

4.3. Probability integral transform histogram for count data

A PIT histogram assesses the predictive quality of a model with respect to calibration. The PIT for a certain region is the value of the predictive cumulative distribution function at the observed count. If the observation was drawn from the predictive distribution—which would be the ideal case—and the predictive distribution is continuous, the PIT values have a standard uniform distribution. As a diagnostic tool, a histogram of the obtained PIT values is plotted and checked for uniformity. If there are deviations from uniformity, forecast failures and model deficiencies might be present. U-shaped histograms indicate underdispersed predictive distributions; hump or inverse U-shaped histograms point to overdispersion (Czado *et al.*, 2009).

In the case of count data, as in the present paper, the predictive distribution is not discrete, and the PITs are no longer uniform under the hypothesis of an ideal forecast. Hence, an adjustment

Table 3. DIC and mean logarithmic score $\overline{\text{LS}}$

Model	$\bar{D}$	p_D	*DIC*	$\overline{LS}$
M1	1712.1	91.3	1803.5	1.011
M2	1622.7	117.8	1740.5	0.979
M3	1674.7	107.0	1781.6	1.001
M4	1516.9	231.6	1748.5	1.002
M5	1645.5	177.5	1823.0	1.052

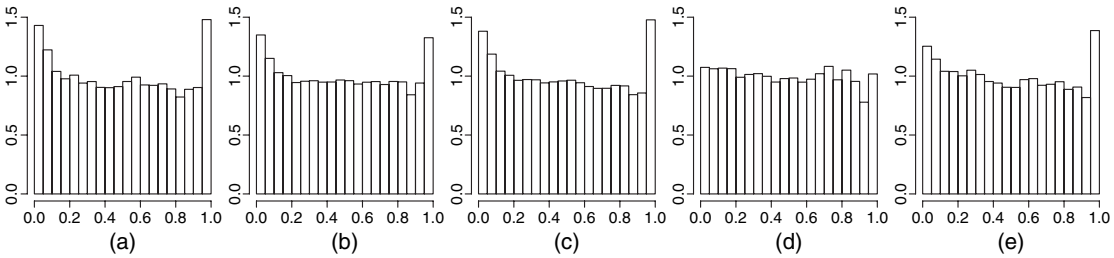


Fig. 1. PIT histograms for models (a) M1, (b) M2, (c) M3, (d) M4 and (e) M5

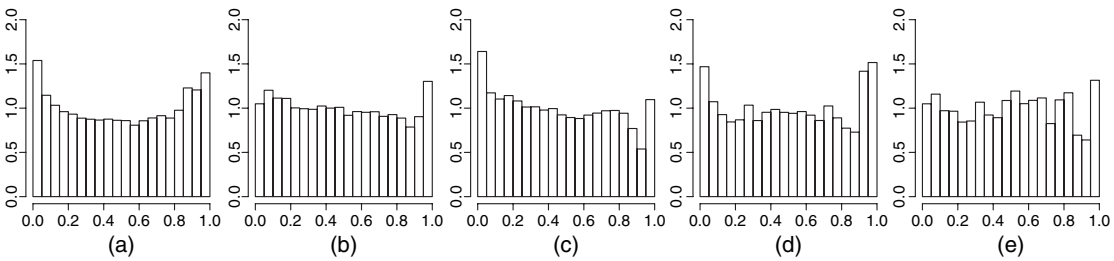


Fig. 2. PIT histograms for model M2, separately for each year: (a) 2003; (b) 2004; (c) 2005; (d) 2006; (e) 2007

is necessary as, for example, suggested by Czado *et al.* (2009). The resulting histogram can be interpreted in the same way as a PIT histogram derived for continuous data.

5. Results by using integrated nested Laplace approximations

All models from Section 2 were fitted to the BVD data by using INLAs (full Laplace approximation). Model choice is conducted in Section 5.1 to find the best model and to determine whether cantonal heterogeneity is present in the data. Some interesting results with regard to under-reporting in the data are presented in Section 5.2.

5.1. Model choice and calibration

The DIC and its components are shown in Table 3 as well as the mean logarithmic scores. For model M4 the fit is best, but it has a large p_D because of the complex dependence structure within the Gaussian field. The best trade-off between model complexity and fit is found for model M2; it has the lowest DIC value. In general, models including a first-order RW time trend are preferred to the analogous model with a second-order RW trend. The mean logarithmic

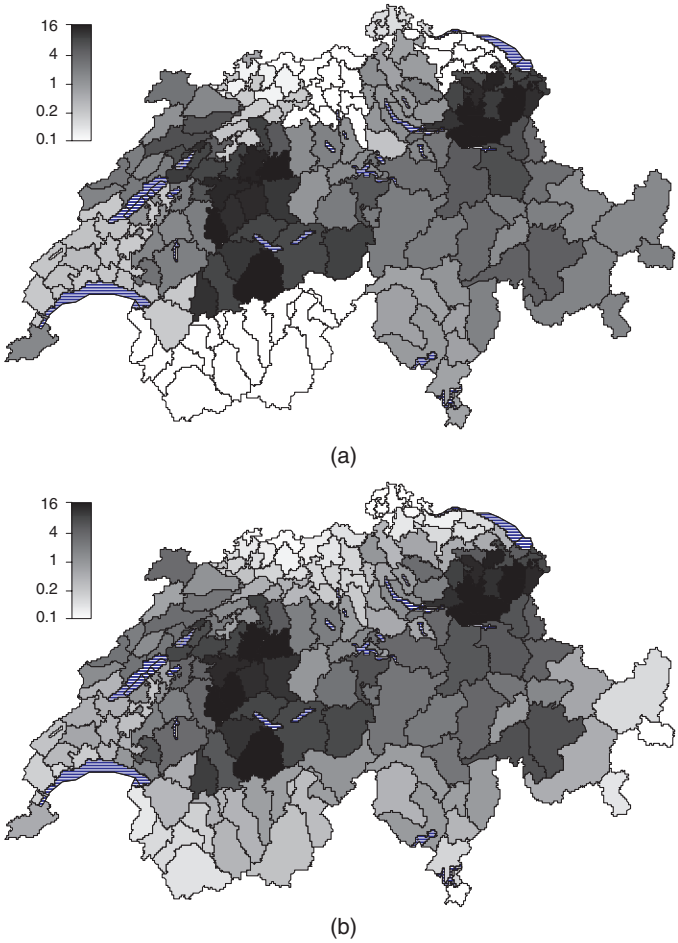


Fig. 3. Fitted relative incidence estimated by (a) model M2 ($\exp(\nu_i + \alpha_{j(i)})$) and (b) model M4 ($\exp(\nu_i + \psi_i)$)

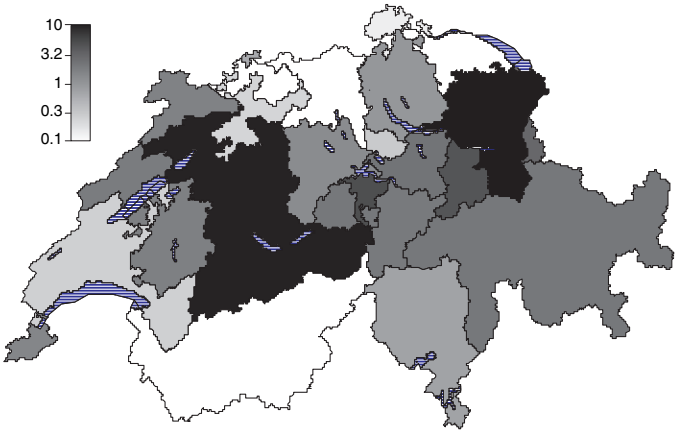


Fig. 4. Cantonal effects α_j on an exponential scale (model M2)

score is also lowest for model M2. Hence, cantonal heterogeneity is present in the data and a first-order RW formulation is most appropriate for the time trend; see Table 2.

To investigate model calibration, PIT histograms for all models are shown in Fig. 1. The histograms for models M1, M2, M3 and M5 are close to uniform except for higher columns at the left-hand and right-hand end of the histograms. This indicates underdispersion of the predictive distribution. The PIT histogram for model M4 is very close to uniformity and, hence, calibration is best for this model.

Since model M2 was considered as the best model by the model choice criteria, its calibration should be examined in further detail. Fig. 2 shows a PIT histogram for model M2 separately for each year. Underdispersed predictions are present particularly for the years 2003, 2005 and 2006. Hence, the underdispersed predictions can be attributed to a poor predictive performance in some years.

5.2. Results for the bovine viral diarrhoea data

The fitted relative spatial incidence for each region ($\nu_i + \alpha_{j(i)}$) for the best model (M2) is shown in Fig. 3(a) on an exponential scale; lakes are indicated as striped areas. The large range of relative incidence (0.1–16) indicates biased case reporting since such large differences cannot be explained from the nature of the disease. We now consider the effects of cantons and regions separately to investigate which one is more pronounced. Plots of the cantonal (α_j) and regional effects (ν_i) on an exponential scale are shown in Figs 4 and 5(a) respectively. The cantonal effect has a larger influence on the total relative incidence than the effect on regional level, which indicates that there is a strong heterogeneity in reporting between cantons. The cantons Berne, St Gallen, Appenzell-Innerrhoden and Appenzell-Ausserrhoden show an increased relative incidence for BVD which is elevated by a factor of 10. The incidence of a reported BVD case is lowest in the cantons Valais, Aargau and Thurgau. This is clear evidence for under-reporting in the data due to different policies of the cantonal authorities. An unstructured spatial heterogeneity between regions is also present; see Fig. 5(a). This map might represent regional differences in disease prevalence. It is also suspected that there are regions with single stockholders who are aware of the disease or have faced financial damage caused by BVD in the past. To investigate what happens if only regional terms are included in the model the incidence fitted by model M4 is plotted in Fig. 3(b). Clearly, cantonal borders are not taken as much into account as in model M2. Cantonal borders are shown in Fig. 4.

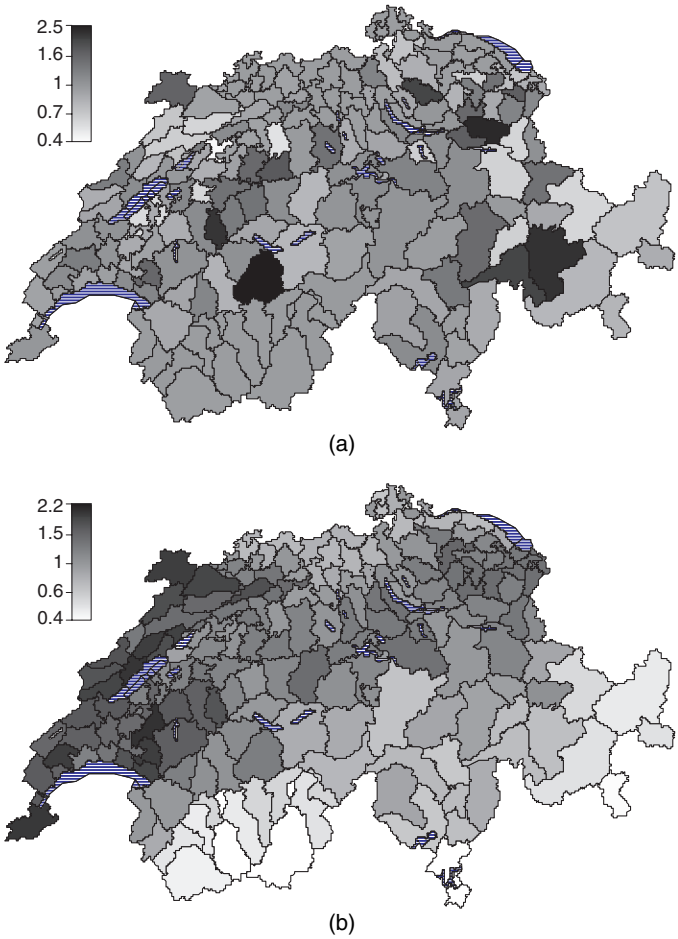


Fig. 5. Regional effects ν_i on an exponential scale (model M2) and relative risk of BVD in 2008

With model M1, the main linear time effect ϕ was estimated as 0.28 with 2.5%- and 97.5%-quantiles of -0.03 and 0.55 respectively, indicating a positive trend. The estimated log-rate for time (including μ) for all models is shown in Fig. 6(b). Models M2 and M4 also show an increasing time trend, but a more pronounced increase in reported cases for the years 2006 and 2007 compared with the three preceding years. This large rise in reported cases can be explained by the increasing amount of information on BVD which was given to stockholders by the Swiss Federal Veterinary Office from the end of 2005.

Estimates of the cantonal time trend $(\mu + \alpha_j + (\phi + \varphi_j)t)$ that were obtained by model M1 are shown in Fig. 6(a). A strong positive differential trend φ_j can be observed for the canton Fribourg and Liechtenstein. In these two areas the rise in reported cases was steeper than on average. The canton Vaud shows a strong negative differential trend. The posterior mean of the correlation is 0.39 with a 95% credible interval of $[-0.07; 0.75]$. This means that a positive correlation between cantonal effect and differential trend is present; cantons with similar relative risk estimates behave fairly similarly over time. Evaluating the estimated interaction effects for the best non-parametric model, M2, a strongly positive trend is found for Fribourg and Liechtenstein in the year 2007. This shows that in these two areas an immediate rise in reported cases in the year 2007 took place, which was even stronger than the mean trend. This is also true

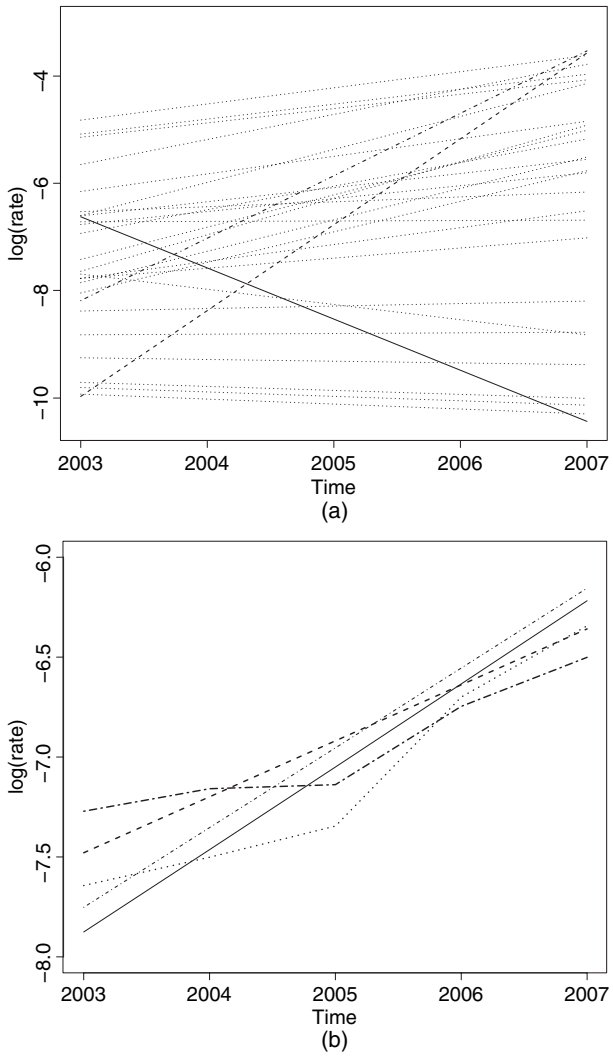


Fig. 6. (a) Linear time trend for each canton (model M1) on a log-scale (- - - - - - -, Fribourg; · · · · · · · ·, Liechtenstein; ———, Vaud; · · · · · · ·, others) and (b) estimated main time trend for all models on a log-scale (including μ) (- - - - - - -, model M1; ·—·—·—·, model M2; · · · · · · ·, model M3; · · · · · · · ·, model M4; ———, model M5)

for the canton Lucerne in the years 2006 and 2007. In contrast, the number of reported cases in the canton Vaud decreased in 2006 and 2007.

To investigate whether the mean herd size per region has an influence on incidence of the disease, an ecological regression was conducted. The logarithm of the mean herd size was included in the linear predictor (2) of model M2 as an explanatory variable. The resulting point estimate is -0.45 with a 95% credible interval of $[-1.26; 0.38]$. Hence, for BVD no clear association between log(mean herd size) per region and incidence of disease could be found.

Swiss cantons vary greatly with respect to size and number of regions. There are cantons which consist of only one region, whereas, for example, the canton Berne is split into 26 regions. This fact has an influence on the fit of the models proposed. Fig. 7 shows the logarithm of the observed proportion and the fitted values of models M1, M2 and M4 for four regions: Fig. 7(a) for the canton Appenzell-Innerrhoden, which is both a region and a canton, Figs 7(b) and 7(c) for two of five regions of the canton Lucerne and Fig. 7(d) for one region in the canton Valais, where

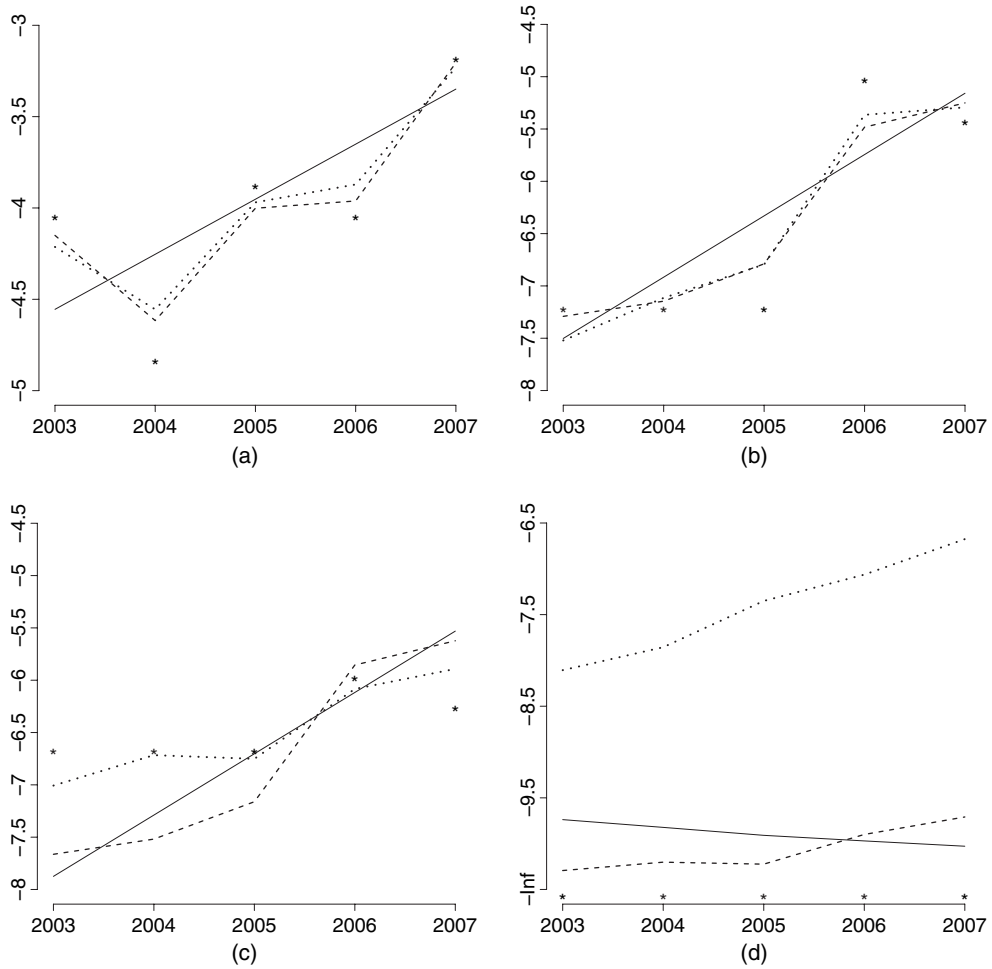


Fig. 7. Logarithm of the observed proportion (∗) and fitted values for models M1 (———), M2 (- - - - - - -) and M4 (· · · · · · ·) for (a) the canton Appenzell-Innerrhoden, (b), (c) two regions in the canton Lucerne and (d) one region in the canton Valais

not a single case of BVD was reported during the whole time period. Here, the observed proportion on the logarithmic scale is equal to $-\infty$. As the behaviour of models M3 and M5 is very similar to that of model M1 in all cases, the results are not included in these figures. In Fig. 7(a) models M2 and M4 behave quite similarly, adjusting very well to the non-linear time trend in Appenzell-Innerrhoden. The situation is different in Figs 7(b) and 7(c): the fit of models M2 and M4 is similar in Fig. 7(b), but very different in Fig. 7(c). Model M4 is, in contrast with the models including a cantonal effect, sensitive to regional departures from a cantonal time trend which is present in the region shown in Fig. 7(c). In models including interactions between time and canton, the shape of the fitted time effect is equal for each region within one canton; just the level of the fitted values can change. This smoothing effect of models M1 and M2 can be noted for larger cantons. In contrast, for some regions models M1 and M2 are more sensible: Fig. 7(d) shows the fit for a region in the canton Valais where not a single case of BVD was reported during the whole time period. As model M4 does not take into account cantonal borders, the incidence of BVD is estimated too high in those regions that are close to the borders of cantons with reported cases.

6. A comparison of integrated nested Laplace approximations and Markov chain Monte Carlo methods

To assess the accuracy of the estimates that were obtained by INLAs, the best model (M2) was analysed by using MCMC methods. Unfortunately, freely available MCMC software like WinBUGS (Lunn *et al.*, 2000) and BayesX (Brezger *et al.*, 2005) does not incorporate linear constraints properly. Instead, so-called 'centring on the fly' is used as an *ad hoc* approach to

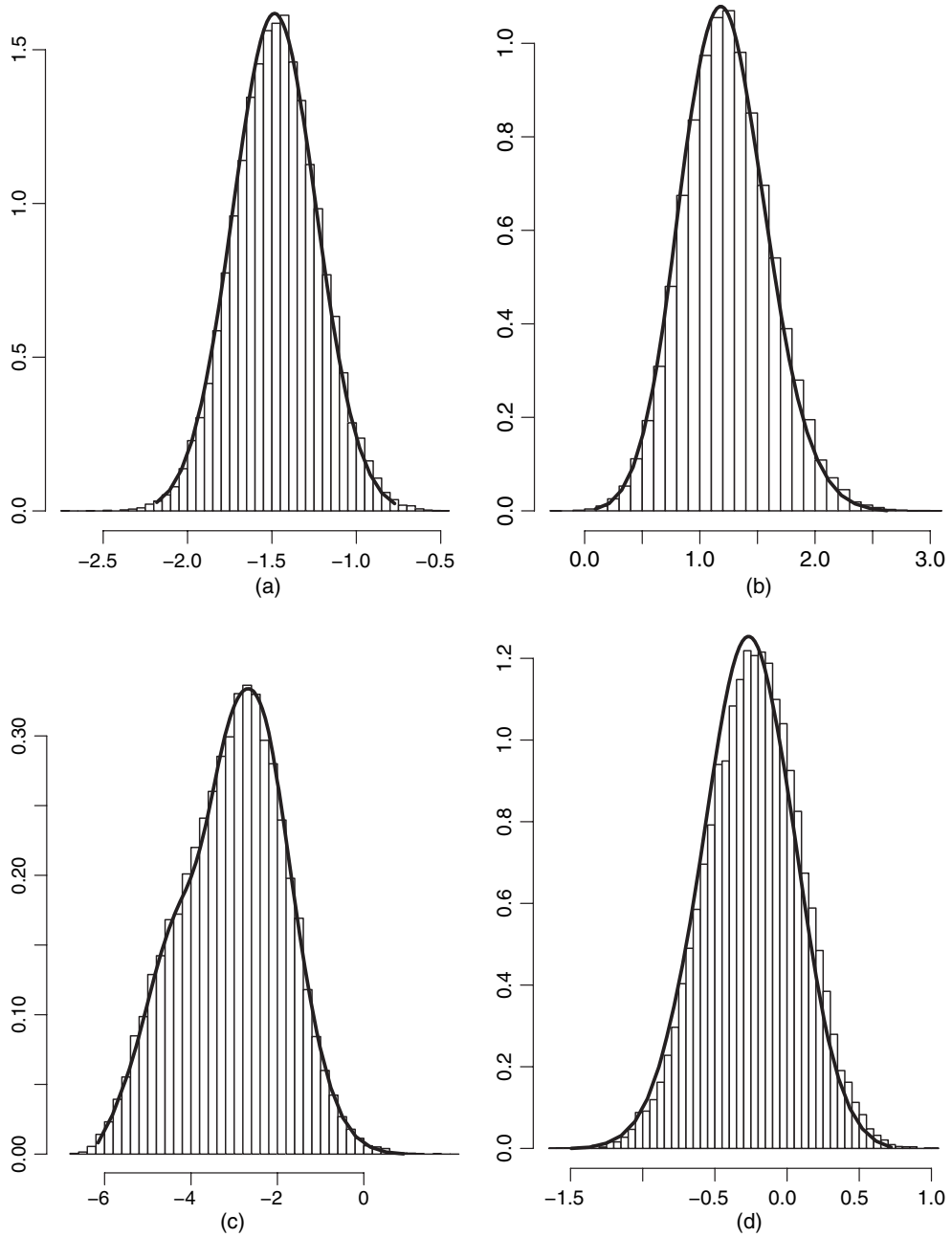


Fig. 8. MCMC histograms and INLA approximations of the posterior marginals for all variances within model M2 (on a log-scale): (a) $\log(\sigma_{\nu}^2)$; (b) $\log(\sigma_{\alpha}^2)$; (c) $\log(\sigma_{\beta}^2)$; (d) $\log(\sigma_{\delta}^2)$

incorporate sum-to-zero constraints. However, it is unclear whether the resulting algorithm has the correct equilibrium distribution. Hence, advanced MCMC routines were implemented by the third author using a low level programming language (C) with correct incorporation of all linear constraints as discussed in Rue and Held (2005). Two different approaches were used to obtain samples of the posterior marginals, namely auxiliary mixture sampling (Frühwirth-Schnatter *et al.*, 2009) and a Metropolis–Hastings algorithm with a proposal constructed by using a second-order Taylor expansion of the log-likelihood (Rue and Held, 2005). To obtain very precise estimates that are suitable for a comparison of MCMC methods and INLAs $n_s = 3\,030\,000$

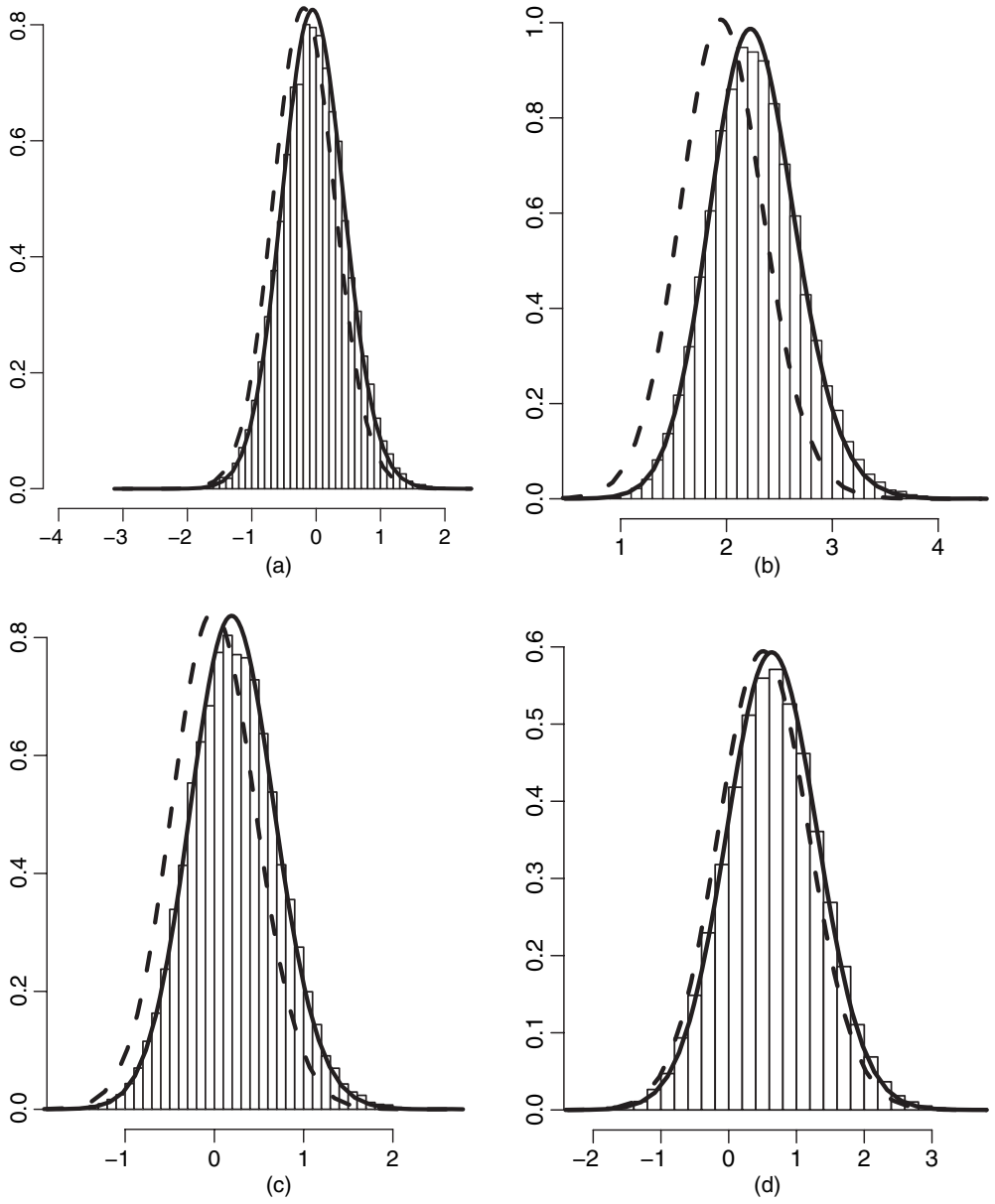


Fig. 9. MCMC histograms and INLA approximations of the posterior marginals of four Swiss cantons (α_j, on a log-scale; the INLA approximations were obtained by using the Gaussian (– – –) and the full Laplace approximation (———)): (a) Zurich; (b) Berne; (c) Lucerne; (d) Uri

samples were drawn by using a thinning of 100 and a burn-in of 30 000. Hence, 30 000 samples were left for an estimation of the posterior quantities. Negligible auto-correlation was left in these samples. The Monte Carlo standard error of the estimates for each component of the latent field and the variances was estimated by using the method of consistent batch means (Jones *et al.*, 2006). We used $\sqrt{n_s}$ to determine the size of batch. The resulting Monte Carlo error estimates were smaller than 0.01 in each case, except for σ^2_β ($\text{se}(\sigma^2_\beta) = 0.43$). Nevertheless, an application of the stopping criterion that was described in Jones *et al.* (2006) confirmed that the length of our MCMC chain is sufficient for this parameter. The results for auxiliary mixture sampling and the Taylor approximation were virtually identical and only results by using the Taylor approximation are shown.

To compare INLA and MCMC methods, histograms of the MCMC samples and the approximations of the posterior marginals by INLA are compared in Figs 8 and 9. For the variance components of model M2 (see equation (2)) these plots are shown in Fig. 8, with variances shown on the log-scale. The MCMC histograms and INLA approximations are virtually identical. For the latent Gaussian field $\mathbf{x}$ MCMC histograms and INLA approximations look virtually identical for all components $(\mu, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta})$, when the simplified or full Laplace approximation are used for the approximation of part 2 in expression (10). Small shifts can be observed for the cantonal components when the Gaussian approximation is used; see Fig. 9. Hence, for an improved approximation the simplified or full Laplace approximation must be used.

Unfortunately, the incorporation of linear constraints slows down the computation by INLAs for the simplified and full Laplace approximation. The computer times for model M2 are 26.79 s (Gaussian approximation), 105.32 s (simplified Laplace approximation) and 233.23 s (full Laplace approximation). Depending on the number of linear constraints and the data, the difference in computer time between the Gaussian and the simplified or full Laplace approximation can be very high. The MCMC sampler produced 246 iterations per second.

The DIC that was obtained with MCMC sampling is $1622.2 + 119.2 = 1741.4$. The DIC that was computed by INLAs (full Laplace approximation) is very close to this value; see Table 3. We also computed the logarithmic score and the PIT histogram from the MCMC samples by using importance sampling (Stern and Cressie, 2000). However, the distribution of the importance weights was heavily skewed and dominated by a small number of extreme values. In such circumstances the estimates are known to be unreliable (Marshall and Spiegelhalter, 2003). In INLAs, the computation of these quantities fails for a few observations, which are indicated by the `inla` program (Martino and Rue, 2009) , but can easily be obtained by rerunning the model without one of these observations in turn. For a detailed comparison of PIT and logarithmic score from INLA and MCMC methods see Held *et al.* (2010).

Another issue which must be addressed is the usability of both approaches. INLAs can easily be run by using R and all output can be processed directly. This is even true for the complex class of spatiotemporal disease mapping models that was introduced in Section 2. In contrast, to use MCMC techniques, complex algorithms must be implemented by hand and care must be taken concerning the samples obtained.

7. Discussion

Regarding the BVD data, model M2 is chosen as the best model by the model choice criteria that were considered. Hence, cantonal heterogeneity is present in the data and the affiliation of a Swiss region to a certain canton highly influences the number of reported BVD cases. Furthermore, a non-parametric formulation of the time trend which points out an immediate rise in reported cases for the years 2006 and 2007 is adequate. This finding gives rise to the hypothesis

that the disease awareness regarding BVD has been rising since 2006 when information on the disease was given to stockholders by the Swiss Federal Veterinary Office. In 2008, a large-scale programme which included testing every cow in Switzerland started, to eradicate this disease by the end of 2011. The estimated relative risk for the disease in 2008 is shown in Fig. 5(b) (Besag *et al.*, 1991). The pattern obtained differs considerably from the pattern that was found in Fig. 3(a). Hence, pronounced under-reporting is present in the analysed case reporting data for the years 2003–2007. Reasons for the cantonal differences in case reporting must be found and the policy makers should think of strategies to prevent them.

Our analysis shows that INLAs are a flexible and useful tool that can be used to fit spatio-temporal models. Furthermore, the results provided can easily be used for data analysis. However, some experience in choosing the most appropriate approximation technique and appropriate settings for the approximation routines is needed. A comparison with results from an MCMC analysis in Section 6 showed that INLA approximations and MCMC histograms are virtually identical for hyperparameters and the components of the latent Gaussian field, if the simplified or the full Laplace approximation is used.

Acknowledgements

Financial support by the Swiss Federal Veterinary Office is gratefully acknowledged. Many thanks go to Sarah Haile for checking the manuscript and to Håvard Rue for the INLA support. The revision has benefited from very helpful comments and suggestions by two reviewers.

References

Assunção, R., Reis, I. and Oliveira, C. (2001) Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statist. Med.*, **20**, 2319–2335.

Banerjee, S., Carlin, B. and Gelfand, A. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman and Hall–CRC Press.

Bernardinelli, L., Clayton, D. and Montomoli, C. (1995a) Bayesian estimates of disease maps: how important are priors? *Statist. Med.*, **14**, 2411–2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. and Ghislandi, M. (1995b) Bayesian analysis of space-time variation in disease risk. *Statist. Med.*, **14**, 2433–2443.

Besag, J., York, J. and Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 1–59.

Breslow, N. and Clayton, D. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Brezger, A., Kneib, T. and Lang, S. (2005) BayesX: analyzing Bayesian structural additive regression models. *J. Statist. Softwr.*, **14**, 1–22.

Clayton, D. (1996) Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 275–301. London: Chapman and Hall.

Czado, C., Gneiting, T. and Held, L. (2009) Predictive model assessment for count data. *Biometrics*, **65**, 1254–1261.

Doherr, M. and Audige, L. (2001) Monitoring and surveillance for rare health-related events: a review from the veterinary perspective. *Phil. Trans. R. Soc. Lond.* B, **356**, 1097–1106.

Frühwirth-Schnatter, S., Frühwirth, R., Held, L. and Rue, H. (2009) Improved auxiliary mixture sampling for hierarchical model of non-Gaussian data. *Statist. Comput.*, **19**, 479–492.

Gneiting, T. and Raftery, A. E. (2007) Strictly proper scoring rules, prediction, and estimation. *J. Am. Statist. Ass.*, **102**, 359–378.

Hedeker, D. and Gibbons, R. D. (2006) *Longitudinal Data Analysis*. Hoboken: Wiley.

Held, L. and Rue, H. (2010) Conditional and intrinsic autoregressions. In *Handbook of Spatial Statistics* (eds A. Gelfand, P. Diggle, M. Fuentes and P. Guttorp). Boca Raton: Chapman and Hall–CRC Press.

Held, L., Schrödle, B. and Rue, H. (2010) Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures—Festschrift in Honour of Ludwig Fahrmeir* (eds T. Kneib and G. Tutz). Heidelberg: Physica.

Jones, G., Haran, M., Caffo, B. and Neath, R. (2006) Fixed-width output analysis for Markov chain Monte Carlo. *J. Am. Statist. Ass.*, **101**, 1537–1547.

Knorr-Held, L. (2000) Bayesian modelling of inseparable space-time variation in disease risk. *Statist. Med.*, **19**, 2555–2567.

Knorr-Held, L. and Rue, H. (2002) On block updating in Markov random field models for disease mapping. *Scand. J. Statist.*, **29**, 597–614.

Lagazio, C., Biggeri, A. and Dreassi, E. (2003) Age-period-cohort models and disease mapping. *Environmetrics*, **14**, 475–490.

Langford, I., Bentham, G. and McDonald, A.-L. (1998) Multi-level modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European community. *Statist. Med.*, **17**, 41–57.

Langford, I. H., Leyland, A. H., Rasbash, J. and Goldstein, H. (1999) Multilevel modelling of the geographical distributions of diseases. *Appl. Statist.*, **48**, 253–268.

Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statist. Comput.*, **10**, 325–337.

Marshall, E. and Spiegelhalter, D. (2003) Approximate cross-validatory predictive checks in disease mapping models. *Statist. Med.*, **22**, 1649–1660.

Martino, S. and Rue, H. (2009) Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: a manual for the `inla` program. Norwegian University for Science and Technology, Trondheim. (Available from `http://www.math.ntnu.no/~hrue/GMRFLib`.)

Natario, I. and Knorr-Held, L. (2003) Non-parametric ecological regression and spatial variation. *Biometr. J.*, **45**, 670–688.

R Development Core Team (2005) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Richardson, S., Abellan, J. and Best, N. (2006) Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire. *Statist. Meth. Med. Res.*, **15**, 385–407.

Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields*. Boca Raton: Chapman and Hall–CRC Press.

Rue, H. and Martino, S. (2007) Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *J. Statist. Planng Inf.*, **137**, 3177–3192.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc.* B, **71**, 319–392.

Schmid, V. and Held, L. (2004) Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics*, **60**, 1034–1042.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.

Stern, H. and Cressie, N. (2000) Posterior predictive model checks for disease mapping models. *Statist. Med.*, **19**, 2377–2397.

Stone, M. (1977) An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Statist. Soc.* B, **39**, 44–47.

Swiss Federal Veterinary Office (2006) BVD im Detail. Swiss Federal Veterinary Office, Berne. (Available from `http://www.bvet.admin.ch/gesundheit_tiere/00286/00288`.)

Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Ass.*, **81**, 82–86.

PAPER VI

Assessing the impact of network data on the spatio-temporal spread of infectious diseases

Birgit Schrödle, Leonhard Held & Håvard Rue

Paper in revision for *Biometrics*.

Assessing the impact of network data on the spatio-temporal spread of infectious diseases

Birgit Schrödle[1], Leonhard Held[1] and Håvard Rue[2]

[1] Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Switzerland
[2] Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim

Networks of moving individuals like traded animals between farms represent a potential risk for the spatio-temporal spread of an infectious disease. To assess this relationship we propose two frameworks, namely parameter- and observation-driven models. We discuss both approaches in the context of uni- and multivariate time series of counts with specific emphasis on the direct inclusion of network data. In contrast to observation-driven models, where previous cases are included directly, the disease incidence in a parameter-driven model is governed by a latent stochastic process. We present ready-to-use software based on integrated nested Laplace approximations for inference in parameter-driven models. The predictive performance of both formulations is assessed using proper scoring rules and a score regression approach. The impact of cattle trade on the spatio-temporal spread of Coxiellosis in Swiss cows, 2004-2009, is finally investigated.

Keywords: Infectious disease counts; INLA; Network data; Observation-driven; Parameter-driven; Spatio-temporal.

1 Introduction

Networks of moving individuals represent a potential risk of disease transmission. As an example, cattle trade serves as a contact network between infected herds and the ease of transportation can result in the spread of a disease even over long distances (Fèvre et al., 2006). Bigras-Poulin et al. (2007) state that contact structures have received litte attention from epidemiologists, although they are very important to understand disease transmission. There are well-known events which illustrate the risk associated with animal trade. Gilbert et al. (2005) quantify the strong association between the movements from infected areas and the breakdown of herds for bovine tuberculosis in Great Britain. Fèvre et al. (2006) report that the foot-and-mouth epidemic in 2001 was spread by animal movement from the North of England to France and the Netherlands. This demonstrates even the risk associated with large-distance movement of animals over national borders.
Observed trade patterns are often analyzed using network analysis, where farms are

treated as nodes and the movements of animals are treated as links (Natale et al., 2009). Such networks consider the direction of trade and try to identify farms, which have many contacts and are, therefore, hot spots. Furthermore, it is of interest to identify predictors related to having a high number of contacts, e.g. seasonal variations (Nöremark et al., 2011) and dependencies on enterprise type and gender (White et al., 2010). Often, the distribution of the contacts among farms is skewed with many farms having only few contacts while others have many (Bigras-Poulin et al., 2007). Furthermore, medium and large farms have a significantly higher movement activity than small farms. Nöremark et al. (2011) assess the ingoing infection chain which counts the number of ingoing contacts through other farms. However, in such analyses the trade network is never directly related to the actually observed cases. To assess the disease dynamics based on the network, computer simulations can be conducted using a Markov chain model (Natale et al., 2009). Such simulations can give hints on the effects of targeted removal of nodes. However, strong assumptions on the network must be made. Various probabilistic models were proposed in the context of the above mentioned foot-and-mouth epidemic. Green et al. (2006) use simulations to analyze the relationships between cattle trade and the temporal and spatial patterns of predetection foot-and-mouth cases. Ferguson et al. (2001) establish a model to simulate the impact of movement restrictions and additional control strategies. Jewell et al. (2009) introduce a Bayesian framework for stochastic transmission models and apply it as well to the foot-and-mouth data.

The importance of travel networks of humans has also been recognized within the last few years. Hufnagel et al. (2004) combine stochastic local infection dynamics among individuals with stochastic transport in a worldwide network to describe the spread of the severe acute respiratory syndrome (SARS). Merler and Ajelli (2010) predict a rapid diffusion of pandemic influenza within Europe because of the high mobility of the population. Paul et al. (2008) assess the spread of influenza in the US by incorporating air traffic information following investigations by Brownstein et al. (2006).

Our aim in this paper is to formulate a spatio-temporal statistical model that can adapt to epidemic outbreaks and directly relate network data to the observed disease cases. In our case study assessing cattle trade, we want to use appropriate model choice criteria to assess, if the assumption of a cattle trade spread mechanism outperforms a purely adjacency-based, local mechanism. We focus on the analysis of so-called aggregated area-level data for I administrative spatial regions over time. The cattle trade network information is available as an unsymmetrical, possibly time-dependent, $I \times I$ matrix containing the absolute number of traded cattle for each pair of regions.

Two different approaches known from the analysis of time series data are proposed in the following. Cox (1981) states a difference between so-called observation- and parameter-driven models. While in observations-driven models past observations or covariates are included directly (e.g. Zeger and Qaqish, 1988; Li, 1994), the dependence between subsequent observations is modelled by a latent stochastic process in parameter-driven models. This approach seems justifiable, if the disease is mostly subclinical.

A parameter-driven model was proposed by Zeger (1988) in a maximum likelihood framework assuming a Poisson distribution for the observed counts. Unfortunately, the integrals contained in the marginal likelihood of such so-called generalized linear mixed models are intractable (Nelson and Leroux, 2006). Inference for similar models in a hierarchical Bayesian setting has been discussed by many authors (Hay and Pettitt, 2001; Gamerman, 1998; Frühwirth-Schnatter and Wagner, 2006). Mugglin et al.

(2002) use a hierarchical approach to model influenza epidemic dynamics in time and space. They assess the influence of first and second order neighbours on the disease spread by a vector autoregressive process. Knorr-Held and Richardson (2003) propose a discrete-time model incorporating area-specific latent time indicators to distinguish endemic from hyperendemic periods within meningococcal disease cases from France. For point-referenced data, Brix and Diggle (2001) propose a hierarchical log-Gaussian Cox process to monitor possible changes in the space-time incidence pattern of gastrointestinal infections and to predict the variation in the latent intensity. Diggle et al. (2004) review both point-referenced and area-level hierarchical models for infectious disease counts observed in discrete time, and discuss their application within on-line surveillance systems.

Within this paper we propose a general, parameter-driven approach to assess different networks of spatio-temporal disease spread. The building block is a vector autoregressive model including appropriate weights. A major emphasis is on the use of integrated nested Laplace approximations (INLA), a recently proposed method for approximate Bayesian inference in latent Gaussian models (Rue et al., 2009). INLA gives very accurate estimates in short computational time. Its usage is straightforward thanks to an available `R` package. A toolbox for the implementation of so-called dynamic models within INLA is described in Ruiz-Cardenas et al. (2010). Dynamic or state-space models are a broad class of parametric models where both, parameter variation and available data information, are described in a probabilistic way. Many hierarchical models for uni- and multivariate time series considered so far in the literature fit into this framework and can be fitted directly with the `INLA` software (`www.r-inla.org`).

Another class of models, which is an additive mixture of parameter- and observation-driven components, has been proposed in Held, Höhle and Hofmann (2005), denoted by H^3 in the following. Here, the process of infection is modelled by a so-called epidemic part by directly including past observations in the region of interest and/or other regions to describe local epidemics (Paul et al., 2008). Additionally, the linear predictor contains an endemic part, which can consist of temporal and seasonal trends. This model perspective is motivated from a branching process with immigration. An important and very attractive advantage of this formulation is that for simple settings maximum likelihood inference can easily be realized by generic optimization routines, for example the function `optim()` in `R` (R Development Core Team, 2005). Relevant datasets and algorithms are available in the `R` package `surveillance`. For multivariate data random intercepts for each region can be included (Paul and Held, 2011); appropriate software is also freely available at `https://r-forge.r-project.org/projects/surveillance`.

The evaluation of one-step-ahead predictive forecasts is an important issue for the proposed problem, since it is of major interest to predict the future incidence associated with cattle trade. An evaluation of probabilistic one-step-ahead predictions derived from the parameter-driven and H^3 model can be done using proper scoring rules (Gneiting and Raftery, 2007; Czado et al., 2009). In this paper we chose the squared error, the logarithmic and the ranked probability score, since they can easily be calculated from the INLA output for the parameter-driven and for the H^3 model (Paul and Held, 2011). Furthermore, to assess the calibration of the predictive distributions we apply a score regression approach based on the Dawid-Sebastiani score (Held et al., 2010).

For illustration, both methodologies are first applied to a univariate time series of

weekly Salmonella agona counts in the UK, 1990-1995. This example is used to clearly state the different properties of observation- and parameter-driven models. The actual data of interest are Coxiellosis cases in Swiss cows from 2004-2009. They are available for 184 Swiss regions and the Principality of Liechtenstein. Coxiellosis is a widespread infectious, bacterial disease which is mostly subclinical. The spread happens from animal to animal by airborne infection. The bacterium can cause an abortion even in a late phase of the pregnancy and is mostly detected by a mandatory screening test due to such an event. Furthermore, Coxiellosis is a zoonosis and can cause, e.g., Q-fever in humans. We will attempt to assess the spatio-temporal spread of the disease caused by cattle trade. Since it is recently mandatory in Switzerland to notify cattle trade, trade data are available for all Swiss regions for 2009. Cattle trade data for earlier years are not available, but experts stated that the trade pattern has not changed for at least the observed time period.
The rest of this paper is organized as follows: Section 2 provides an introduction to the INLA methodology and to proper scoring rules as building blocks for approximate Bayesian inference in parameter-driven models and model choice. Section 3 discusses univariate Salmonella agona counts as an illustrative example. Section 4 deals with multivariate time series models, their extension and implementation within INLA. Furthermore, the proposed models are applied to the Swiss Coxiellosis data. We close with a discussion of the findings in Section 5. The explicit `R` code is given in the Web Appendix A.

2 INLA, computation and model choice

Integrated nested Laplace approximations are a recently proposed method for approximate Bayesian inference in latent Gaussian models (Rue et al., 2009). The major advantage of INLA in comparison to the widely used MCMC algorithms is that it gives very precise estimates in short computational time. The posterior marginals are approximated directly by INLA and no Monte Carlo samples have to be drawn. The usage of MCMC algorithms for generalized dynamic models such as proposed for time series data is only feasible using data augmentation and requires complex sampling schemes to guarantee efficiency (Gamerman, 1998; Frühwirth-Schnatter et al., 2009). The inclusion of random effects for multivariate spatio-temporal data might require an elaborate re-parametrization of the model (Chib et al., 1998). In contrast, INLA can be run within a user-friendly `R` environment (R Development Core Team, 2005) using generic functions; all code is freely available from `www.r-inla.org` and can easily be installed. All analyses within this paper were run using the `INLA` version built on February 3rd, 2011. The `R` code for all examples in Sections 3.2 and 4.2 is provided in Web Appendix A. A detailed description of the INLA methodology is not given here, since this was done by various authors and for different applications. See, for example, Schrödle et al. (2011) in the context of spatio-temporal disease mapping models and Riebler et al. (2010) for correlated multivariate age-period-cohort models. It is also shown in both papers that there is excellent agreement between INLA and MCMC results.
A strength of INLA is that measures for model choice can easily be derived from the `INLA` output. For time series the predictive performance with respect to one-step-ahead predictions is of major interest. One-step-ahead predictions are obtained by successively leaving out the last observation, re-fitting the model and predicting the next count. The performance of probabilistic predictions can be evaluated using proper

scoring rules (Gneiting and Raftery, 2007). Scoring rules assign a numerical score based on the predictive distribution and the actually observed count to each prediction and facilitate model comparison and selection. They address sharpness - the concentration of the predictive distribution - as well as calibration - the statistical consistency between the predicted and the later observed probability distributions (Paul and Held, 2011). The concept is elaborated for count data by Czado et al. (2009). Prominent examples of such scores are the squared error score (SES), the logarithmic score (logS) and the ranked probability score (RPS) (Paul and Held, 2011). Each score has specific properties and Czado et al. (2009) recommend to investigate several scores in applications. The scores used in this paper are defined as

$$\begin{aligned}
\text{SES}(P, y) &= (y - \mu_P)^2 \\
\text{logS}(P, y) &= -\log(P(Y = y)) \\
\text{RPS}(P, y) &= \textstyle\sum_{k=0}^{\infty} (P(Y \leq k) - \mathbf{1}(y \leq k))^2.
\end{aligned}$$

Here, P is the predictive probability distribution, μ_P is its first moment and y the truly observed value. If scores are calculated for several predictions (e.g. for several timepoints or regions), the mean score is computed and compared. The smaller the mean score, the better the predictive performance of the model. A formal comparison of the mean scores of two competing models can be conducted by a Monte Carlo permutation test for paired individual scores (Paul and Held, 2011). The proposed scores are used to compare parameter-driven (PM) and H^3 models in Sections 3.2 and 4.2.

To compare the calibration of two predictive distributions, a score regression approach based on the Dawid-Sebastiani score (DSS) can be used (Held et al., 2010). The DSS is based on the first two moments of the predictive distribution (μ_P, σ_P^2). Held et al. (2010) state that if these two moments match the corresponding moments of the data-generating distribution (ideal forecast), the expectation of the DSS depends only on the log-predictive standard deviation. The DSS is defined as

$$\text{DSS}(P, y) = \frac{1}{2} \cdot (\log(\sigma_P^2) + ((y - \mu_P)/\sigma_P)^2).$$

The expectation is $\frac{1}{2} + \log(\sigma_P)$ for an ideal forecast. The smallest possible DSS is $\log(\sigma_P)$. Hence, we regress the individual scores on the logarithm of σ_P by a standard linear regression model

$$\text{DSS}_i = \kappa + \tau \cdot \log(\sigma_{P,i}) + \epsilon_i.$$

For an ideal forecast $\kappa = \kappa_0 = \frac{1}{2}$ and $\tau = \tau_0 = 1$. To assess the null hypothesis $H_0 : \ \kappa = \kappa_0$ and $\tau = \tau_0$ we can construct an asymptotic χ^2-test with 2 degrees of freedom using an appropriate test statistic (Held et al., 2010).

3 Univariate time series

3.1 Statistical modelling

In parameter-driven models it is assumed that an unobserved autocorrelated mechanism drives the infection process (Zeger, 1988). Consider a model

$$Y_t \sim \text{Po}\big(\exp(\eta_t)\big) \quad \text{with } t = 1, ..., T$$

Table 1: Mean squared error score ($\overline{\text{SES}}$), mean logarithmic sore ($\overline{\text{logS}}$) and mean ranked probability score ($\overline{\text{RPS}}$) for all PM and H^3 models for the Salmonella agona data. The scores are obtained by averaging the scores of one-step-ahead predictions for the last 100 observations. A negative binomial distribution was assumed for the H^3 models.

Model	Autoregr. term	Seasonal term	$\overline{\text{SES}}$		$\overline{\text{logS}}$		$\overline{\text{RPS}}$	
			PM	H^3	PM	H^3	PM	H^3
1	λ	–	4.105	4.249	2.025	2.059	1.116	1.148
2	–	$\sin(\omega t)/\cos(\omega t)$	4.558	4.550	2.173	2.115	1.167	1.183
3	λ	$\sin(\omega t)/\cos(\omega t)$	4.012	4.084	2.011	2.045	1.106	1.126

for a univariate time series of counts, which is formulated in a hierarchical Bayesian framework. The disease risk is modelled by the linear predictor η_t:

$$\begin{aligned} &\text{Stage 1:} \quad \eta_t = \alpha + \zeta_t \\ &\text{Stage 2:} \quad \zeta_t = \lambda \cdot \zeta_{t-1} + \epsilon_t. \end{aligned} \tag{1}$$

This is a so-called dynamic model, where Stage 2 describes the evolution of $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_T)^T$. The vector $\boldsymbol{\zeta}$ forms an autoregressive process of first order (AR(1)), where $\zeta_1 \sim N(0, \sigma_\zeta^2/(1-\lambda^2))$. The errors $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_T)^T$ are assumed to be iid normally distributed with variance σ_ζ^2. The model is stationary, if $|\lambda| < 1$. The parameter α on Stage 1 represents an intercept. In a full Bayesian setting hyperpriors for σ_ζ^2 and λ have to be defined. For convenience, the parameter λ is transformed to

$$\lambda^* = \text{logit}((\lambda + 1)/2)$$

using Fisher's z-transformation; λ^* then takes values over the whole real line.
Time series of infectious disease counts often show a seasonal pattern which can be accounted for by expanding model (1) on Stage 1 to

$$\eta_t = \alpha + \zeta_t + \beta \cdot t + \sum_{s=1}^{S} (\gamma_s \cdot \sin(\omega_s t) + \delta_s \cdot \cos(\omega_s t)). \tag{2}$$

Here, S is the number of harmonics to include and ω_s are the respective Fourier frequencies, for example $\omega_s = 2s\pi/52$ for weekly data.
Several variants of this model will be compared to the H^3 models introduced by Held et al. (2005). Here, the idea is to include the number of cases y_{t-1} in the past directly in the linear predictor to adjust for local epidemics. It is assumed that the counts are Poisson or negative binomially distributed with mean

$$\mu_t = \lambda \cdot y_{t-1} + \exp(\eta_t).$$

The term η_t can be defined equivalently to Stage 1 of the parameter-driven models (1) or (2), but without $\boldsymbol{\zeta}$. This formulation is motivated by a branching process model with immigration and stationary for $0 < \lambda < 1$ (Guttorp, 1995). Given stationarity, the endemic incidence is persistent with a stable temporal, perhaps seasonal pattern. The epidemic incidence will break out occasionally and eventually burn out (Held et al., 2005).

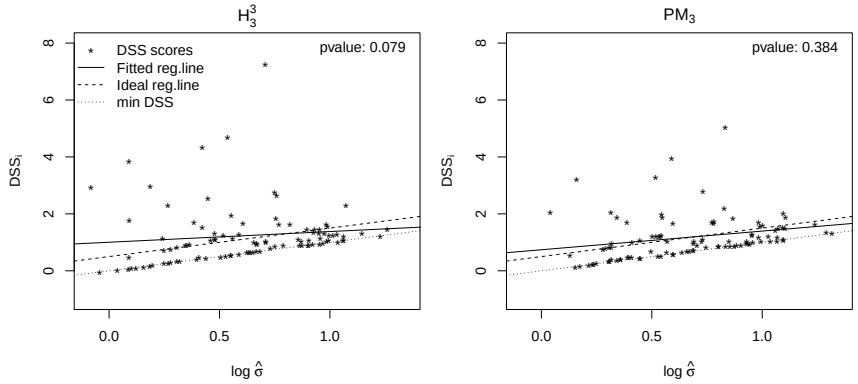


Figure 1: Individual DSS scores, fitted and ideal regression line and minimum DSS score are shown for models H_3^3 (left) and PM_3 (right) based on 100 one-step-ahead forecasts. Additionally, the p-value for a test of miscalibration of the predictive distribution is displayed.

3.2 Salmonella agona in the UK, 1990-1995

As an illustrative example we consider weekly reported cases of Salmonella agona in the UK, 1990-1995. Different models with regard to the autoregressive and seasonal components were fitted, see Table 1. For all seasonal models of type (2) the number of harmonics $S = 1$ was sufficient.

Hyperpriors for the autoregressive and variance components must be defined for the parameter-driven models. We chose a N(0,5) prior for the z-transformed λ^* (Paul et al., 2010). Note that in the INLA setting the hyperprior with regard to the variance of an AR(1) process is imposed on the unconditional variance $\sigma_\zeta^2/(1-\lambda^2)$. We chose a highly dispersed IGa(0.1,0.001) distribution (Frühwirth-Schnatter and Wagner, 2006). A non-informative N(0,10^3) distribution is used as hyperprior for the intercept α and all fixed effects as an INLA default.

To assess the predictive performance of all models, mean scores of one-step-ahead predictions for the last 100 observations were calculated. Note that the PM models are more complex due to their hierarchical structure. If we count the absolute number of (time-independent) parameters, the PM models have one more than the H^3 models (assuming a Poisson distribution), namely the smoothing variance σ_ζ^2. Hence, to make for a fair comparison, the H^3 models were fitted assuming a negative binomial distribution for the response, which contains an additional overdispersion parameter. Nevertheless, the absolute number of parameters does not reflect the true complexity of a hierarchical Bayesian model. An appropriate measure is the effective number of parameters p_D proposed by Spiegelhalter et al. (2002). It is calculated using the deviance of the model and depends on the data. For model PM_3 it grows from around 40 (if the model is fitted to the first 100 observations only) to around 65 (if the model is fitted to all 312 observations). Although the absolute and effective number of parameters cannot be directly compared, this is evidence that the hierarchical model is more complex.

The best model in terms of $\overline{\text{SES}}$, $\overline{\text{logS}}$ and $\overline{\text{RPS}}$ is model PM_3. Model H_3^3 performs best among the H^3 models, but all scores are larger than for the respective PM model. Hence, in this example the predictive performance of the PM models is slightly better.

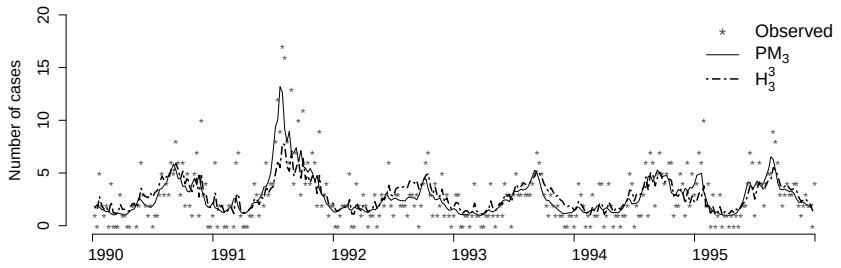


Figure 2: Observed and fitted number of Salmonella agona cases. Fitted number of cases are obtained from the best H^3 model (H_3^3, dashed and dotted line) and the respective parameter-driven model (PM_3, solid line).

However, note that the observational distribution (i.e. negative binomial distribution) is used as predictive distribution for the H^3 models by plugging in the obtained estimates. This approach ignores parameter uncertainty (Paul and Held, 2011). For the INLA results, the full posterior predictive probability distribution for each forecast is approximated directly when re-running the model. As a consequence, the plug-in predictive distribution might be more narrow than the directly approximated predictive INLA distribution. To asses the impact of these two strategies on the calibration of the predictive distribution, we use a score regression approach based on the DSS. The obtained individual DSS scores for 100 one-step-ahead forecasts and the fitted and ideal regression line are shown in Figure 1 for models H_3^3 and PM_3. The estimated regression coefficients are $\hat{\kappa} = 1.00$ and $\hat{\tau} = 0.38$ (H_3^3) and $\hat{\kappa} = 0.74$ and $\hat{\tau} = 0.66$ (PM_3), respectively. The p-values for the test of miscalibration are 0.079 (H_3^3) and 0.384 (PM_3). Hence, calibration is better for the parameter-driven model, but the test detects no significant miscalibration of the predictive distribution for both models.
Figure 2 shows the observed and fitted values for models H_3^3 and PM_3. Apparently, the latent model is more able to adjust for the epidemic peaks. Furthermore, the resulting curve for the parameter-driven model is more smooth.

4 Multivariate time series

4.1 Statistical modelling

Multivariate time series of counts can arise in different contexts, e.g. if time series are available for different age groups, pathogens or geographic regions. We concentrate on spatio-temporal data, but the concepts can also be used in different contexts. Assume that data in region i at time t are available and distributed as

$$Y_{it} \sim \text{Po}\big(m_{it} \cdot \exp(\eta_{it})\big) \quad \text{with } t = 1, ..., T \text{ and } i = 1, ..., I. \tag{3}$$

The quantity m_{it} is an offset that adjusts for possibly different numbers of exposed individuals in each region and time. A multivariate dynamic model can be written in

analogy to model (1) using two stages

$$\begin{array}{ll} \text{Stage 1:} & \eta_{it} = \alpha + \zeta_{it} \\ \text{Stage 2:} & \zeta_{it} = \lambda \cdot \zeta_{i,t-1} + \epsilon_{it}. \end{array} \tag{4}$$

Here, identical replicates of the AR(1) process $\boldsymbol{\zeta}_i = (\zeta_{i,1}, ..., \zeta_{i,T})^T$ for each region i govern the latent evolution. Such a process is often called a Gaussian multivariate or vector autoregressive process and can alternatively be written as

$$\boldsymbol{\zeta}_t = \Omega \cdot \boldsymbol{\zeta}_{t-1} + \boldsymbol{\epsilon}_t, \tag{5}$$

where $\boldsymbol{\zeta}_t = (\zeta_{1,t}, ..., \zeta_{I,t})^T$ and $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, ..., \epsilon_{I,t})^T$ (Harvey, 1981, 1989; Lütkepohl, 2005). The autoregressive coefficient matrix Ω has dimension $I \times I$. In model (4) the matrix Ω is simply diagonal with entries equal to λ.
In order to model the latent spatial spread of the disease, a weighted sum of the past states in other regions j than the region of interest (i) can be included in model (4) on Stage 2:

$$\begin{array}{ll} \text{Stage 1:} & \eta_{it} = \alpha + \zeta_{it} \\ \text{Stage 2:} & \zeta_{it} = \lambda \cdot \zeta_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot \zeta_{j,t-1} + \epsilon_{it}. \end{array} \tag{6}$$

Different choices for the network and, hence, relevant regions j and the respective weights w_{ji} are possible, see also Mugglin et al. (2002). As a first idea all neighbouring regions $i \sim j$ are considered in the ρ term to describe a purely adjacency-driven, local disease spread; in this case $w_{ji} = 1$ for all $i \sim j$ and 0 otherwise. In Section 4.2 the number of traded cattle from region j to region i (and appropriate transformations of this quantity) are used as weights w_{ji} in order to asses the spatio-temporal spread of a cow disease by cattle trade. Model (6) can also be formulated in the form (5), but now Ω also has off-diagonal elements $\rho \cdot w_{ji}$. For the neighbourhood weights the matrix Ω is symmetric. However, it is not symmetric for the cattle trade weights, since the number of cattle traded from region j to region i can be different from the number of cattle traded from region i to region j.
We assume throughout that the cattle trade weights are constant over time, since no time-dependent weights are available for the case study. Since cattle trade also takes place within a region (w_{ii}), we can include this quantity as a fixed weight into the first component on Stage 2, resulting in $\lambda \cdot w_{ii} \cdot \zeta_{i,t-1}$. However, this did not improve the models considered in the case study and is therefore neglected.
One can show that a vector autoregressive AR(1) process is stationary, if all eigenvalues of Ω are less than one in absolute value (Lütkepohl, 2005). However, in practice the region of parameter stationarity cannot be easily derived from this condition (Sun and Ni, 2004). In analogy to the univariate problem, Mugglin et al. (2002) restrict all parameters in the autoregressive coefficient matrix to be in a range of $[-1; 1]$ and assign a Gaussian prior, but, apparently, this transformation does not guarantee stationarity of the multivariate AR(1). For this condition to hold the time series in all regions have to be jointly stationary (Harvey, 1981). Sun and Ni (2004) argue that in the case of no strong prior information the use of reference or noninformative priors is desirable. However, finding suitable reference priors for vector-autoregressive models is almost intractable. For our case study in Section 4.2 we conducted a sensitivity analysis using both Gaussian priors defined on a Fisher z-transformed and unrestricted parameter space.
If the data vary seasonally, a term similar to (2) could be included on Stage 1 of, e.g., model (4). For spatio-temporal data it is also possible to include random components

for each region i on Stage 1 of models (4) and (6), if the disease spread is more of an endemic than epidemic nature. These $\boldsymbol{\nu} = (\nu_1, ..., \nu_I)^T$ random components can be iid normally distributed with mean 0 and variance σ_ν^2. As an alternative an intrinsic conditionally autoregressive (ICAR) effect $\boldsymbol{\psi} = (\psi_1, ..., \psi_I)^T$ can be assumed, which takes into account the incidence in neighbouring regions (Rue and Held, 2005).
The dynamic model (4) can be fitted in INLA using already implemented standard model specifications, see Web Appendix A. For inference of a system like (6) within INLA an augmented model must be coded, where the column of actual observations is merged with a column of pseudo-observations. These pseudo-observations are 0's obtained from equating the evolution equation on Stage 2 to zero. Hence, the resulting response matrix contains two columns, namely

$$\begin{bmatrix} y_{11} & \text{NA} \\ \vdots & \vdots \\ y_{I,T} & \text{NA} \\ -- & -- \\ \text{NA} & 0 \\ \vdots & \vdots \\ \text{NA} & 0 \end{bmatrix}$$

where lines $(1, ..., IT)$ correspond to the actual observations and lines $(IT + 1, ..., 2IT)$ correspond to the pseudo-observations. Different likelihoods are assumed for the observed counts and pseudo-observations (here: Poisson and Gaussian distribution). The `R` code for the case study in Section 4.2 is given in Web Appendix A. In the augmented model the parameters λ and ρ are treated as so-called scaling factors for the respective realizations of the process $\boldsymbol{\zeta}$. A detailed description of the provided toolbox and several similar examples are given in Ruiz-Cardenas et al. (2010) and on `http://www.r-inla.org/models/tools`.
The idea of the proposed models is very similar to H^3 models for multivariate time series (Paul and Held, 2011). In the easiest case, the counts y_{it} are Poisson or negative binomially distributed with mean

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_{it} \cdot \exp(\eta_{it}). \tag{7}$$

An inclusion of weights w_{ii} is possible, but currently not implemented in the available software for H^3 models. Analogous to the univariate case this model can be written as a multivariate branching process with immigration

$$\boldsymbol{\mu}_t = \Lambda \cdot \boldsymbol{y}_{t-1} + \boldsymbol{m}_t \cdot \exp(\boldsymbol{\eta}_t)$$

with suitable defined column vectors $\boldsymbol{\mu}_t$, $\boldsymbol{y}_{t-1}$, $\boldsymbol{m}_t$ and $\boldsymbol{\eta}_t$. The matrix Λ contains the λ's as diagonal and the $\rho \cdot w_{ji}$'s as off-diagonal elements. If the largest eigenvalue of Λ is smaller than unity, the process is ergodic (Held et al., 2005). In analogy to the parameter-driven model the term $\boldsymbol{\eta}_t$ can contain temporal and seasonal components. Furthermore, Paul and Held (2011) discuss the inclusion of regional random effects of iid ($\boldsymbol{\nu}$) or ICAR ($\boldsymbol{\psi}$) type into $\boldsymbol{\eta}_t$. In their framework the autoregressive parameters λ and ρ could also be modelled as random effects and vary from region to region. However, this is not possible in the INLA setting.

Table 2: Total number of pairs of regions and number of pairs with no cattle trade. Furthermore, summary statistics for pairs of regions with cattle trade are displayed.

# of pairs	# 0's	Min	1st Qu	Median	Mean	3rd Qu	Max
34039	17060	1	3	11	49	37	4768

4.2 Coxiellosis in Swiss cows, 2004-2009

The methodology is applied to data on Coxiellosis incidence on Swiss farms. As noted before, the data are available from 2004 to 2009 for 184 Swiss regions and the Principality of Liechtenstein. A herd is denoted a case, if at least one diseased animal was detected. The number of herds per region is included as an offset m_i in equation (3).
Coxiellosis incidence is quite low, with a mean number of 47 cases per year for the observed time period. Hence, the data were yearly aggregated. Such a coarse aggregation can be justified by the fact that the date of confirmation of a case by a laboratory, which is available for our case study, does not directly coincide with the date of an infection. Often the animals have been infected some time before, but the disease is not detected until an abortion takes place. Hence, reporting delay and underreporting are likely to be present. Furthermore, the time order of events can not exactly be reconstructed. To ensure that the temporal aggregation does not dilute any important facts, we re-ran all analyses discussed below using half-yearly and quarterly aggregated data, but the conclusions did not change. However, not all parameter-driven models did converge for half-yearly and quarterly data. As a consequence of the yearly aggregation, seasonality is not modelled for this case study.
A first question to be answered is, if any spatial spread of the disease along a network takes place. Hence, we compare models of type (4) (no spatial spread) with models of type (6) (spatial spread) using predictive scores. Furthermore, different spread mechanisms, namely adjacency-based weights (model index 2) and cattle trade weights, are compared.
Descriptive statistics for the entries of the cattle trade matrix can be found in Table 2. There is no cattle trade between about half of all possible pairs of regions. For those pairs of regions with cattle trade the quantity is very skewed. Hence, a transformation of the cattle trade counts (CT) might be useful. Ideally, a data-driven transformation $f(CT)$ is preferred, but such an algorithm is not yet available. So we compared different choices (see Table 3, model index 3 to 6) based on the one-step-ahead forecast scores, which are discussed below.
Note that hyperpriors for all variance and autoregressive components must be defined for the PM models. For model PM_1, a z-transformed λ^* is used in INLA as a default. The three proposed hyperpriors in Paul et al. (2010) (N(0,1.25), N(0,2.5), N(0,5)) did not result in different posterior marginals and we chose an N(0,5) for this analysis. In analogy to the univariate case a highly dispersed IGa(0.1,0.001) prior is chosen for the unconditional variance of the AR(1) process. For models of type (6), a sensitivity analysis concerning the hyperprior choice for λ and ρ was conducted. As discussed above, we use adjacency-based and cattle trade weights. To guarantee that the chosen hyperpriors mean the same for all weight schemes, the (transformed) cattle trade numbers were standardized in such a way that they are 1 on average. As first and second hyperprior option we chose a N(0,0.25) and N(0,1) prior for λ and ρ. As third

Table 3: Mean squared error score ($\overline{\text{SES}}$), mean logarithmic score ($\overline{\text{logS}}$) and mean ranked probability score ($\overline{\text{RPS}}$) for all PM and H^3 models for the Coxiellosis data. The scores are obtained by averaging over the scores of a one-step-ahead prediction for 2009 for all 185 regions.

		$\overline{\text{SES}}$		$\overline{\text{logS}}$		$\overline{\text{RPS}}$	
Model index	w_{ji}	PM	H^3	PM	H^3	PM	H^3
1	$-$	0.611	0.645	0.583	0.624	0.239	0.257
2	$i \sim j$	0.505	0.603	0.547	0.593	0.218	0.246
3	CT_{ji}	0.494	0.527	0.549	0.590	0.214	0.236
4	CT_{ji}/n_j	0.518	0.540	0.554	0.583	0.218	0.234
5	$\sqrt{CT_{ji}}$	0.644	0.638	0.557	0.619	0.217	0.255
6	$\log(CT_{ji}+1)$	0.750	0.645	0.575	0.624	0.230	0.257
7	$(CT_{ji}/n_j) \cdot b_j$	**0.444**	**0.527**	**0.549**	**0.578**	**0.212**	**0.232**

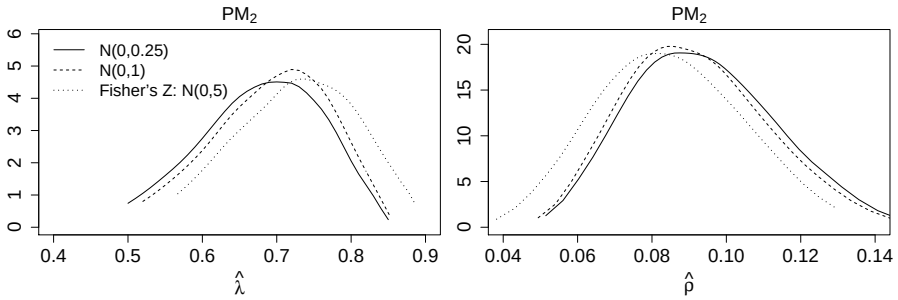


Figure 3: Posterior marginals of λ and ρ for model PM_2 (neighbourhood weights) for three different hyperpriors: an N(0,0.25) (solid) and an N(0,1) (dashed) prior for λ and ρ and an N(0,5) prior for a Fisher's z-transformed λ^* and ρ^* (dotted).

to fifth option we chose the three above mentioned priors for a z-transformed λ^* and ρ^*. Small differences in the posterior marginals can be detected for all five hyperpriors. This is probably due to the sparseness of the data. As an example, the results for prior options 1 to 3 for model PM_2 are shown in Figure 3. Prior options 4 and 5 are omitted from the plot, since their results are very similar to those of option 3. The z-transformed priors resulted in slightly better predictive scores than the unrestricted priors. However, small differences in the hyperparameter estimates do not affect the predictive scores in such a way that the following model comparison is falsified. To guarantee a fair comparison with models of type (4), all the following results are calculated using a z-transformed N(0,5) prior.

Following Paul et al. (2008), we use the absolute (CT_{ji}) and relative number of traded cattle (CT_{ji}/n_j) as weights in model 3 and 4, respectively, see Table 3; here n_j denotes the total number of cattle in region j. Since the cattle trade counts are skewed, a square root ($\sqrt{CT_{ji}}$) and log transformation ($\log(CT_{ji}+1)$) might also be reasonable (models 5 and 6). Models 1 to 6 are fitted using a PM and H^3 approach. A negative binomial distribution was assumed for the H^3 models. To assess the predictive performance of

a model, a multivariate one-step-ahead prediction for all regions for 2009 is made and the mean score over all 185 regions is computed. For the PM models it turns out that the best overall predictive performance taking into account all three scores is obtained using the absolute cattle trade counts as weights (model PM_3). Note that the absolute cattle trade counts also slightly outperform the adjacency-based weights. Among the H^3 models the relative cattle trade performs slightly better than the absolute CT counts for the logarithmic and ranked probability score. The respective scores of model H_4^3 exhibit a better predictive performance than for the neighbourhood weights (H_2^3). The models without spatial spread (PM_1, H_1^3) are clearly outperformed.

An extension of the above discussed models is examined. The mean herd size per region might be a risk factor for the disease spread by cattle trade across regions. The larger a herd, the larger is the risk for at least one subclinically diseased animal. Furthermore, Nöremark et al. (2011) identified herd size as a significant predictor of increased cattle trade activity. In Switzerland, the mean herd size per region b_j varies between about 10 and 60 animals. Hence, we estimated a model with $(CT_{ji}/n_j) \cdot b_j$ as weight w_{ji}, to asses the effect of the mean herd size in the origin trade region j multiplied by the relative cattle trade on the spatial spread of the disease (model PM_7). For models PM_7 and H_7^3 the scores even slightly improved in comparison to models PM_3 and H_4^3.

The variation of the mean scores across different models is generally stronger for the SES than for the logS and RPS, particularly for the PM models. Note that the three scores have different properties. The logS and RPS depend on the full predictive distribution, while the SES depends only on its first moment. The large differences in mean SES between some of the models are caused by only a few observations with different predictive means, to which the mean score reacts severely. In contrast, the logS and RPS appear to be more robust, because the whole predictive distribution is taken into account. The RPS varies less than the logS, which confirms that it is more robust to slight changes in the predictive distribution (Gneiting and Raftery, 2007).

To finally answer the question, if the best transformed cattle trade weights outperform the local adjacency-based weights, we conducted a permutation test for each score and model type. However, the p-values of the permutation tests of PM_2 vs. PM_7 ($p(\overline{\text{SES}}) = 0.48$, $p(\overline{\text{logS}}) = 0.77$, $p(\overline{\text{RPS}}) = 0.45$) and H_2^3 vs. H_7^3 ($p(\overline{\text{SES}}) = 0.18$, $p(\overline{\text{logS}}) = 0.19$, $p(\overline{\text{RPS}}) = 0.06$) do not exhibit a significant difference. But still, the transformed cattle trade weights are at least as good as the adjacency-based weights for the PM and H^3 framework. The differences are larger for the H^3 models.

A further criterion as regards the comparison of the best cattle trade and adjacency-based weights are the p-values of the score regression test for calibration. The respective p-values for the PM and H_3 models are 0.518 (PM_2) vs. 0.738 (PM_7) and 0.424 (H_2^3) vs. 0.571 (H_7^3). Hence, the calibration of the predictive distribution is slightly better for the best cattle trade weights in both modelling frameworks.

In the following the predictive performance of the PM and H^3 models will be compared. All mean scores obtained for the PM models are smaller than for the respective H^3 model. A permutation test comparing the best PM and H^3 models PM_7 and H_7^3 gives the following p-values: $p(\overline{\text{SES}}) = 0.50$, $p(\overline{\text{logS}}) = 0.02$ and $p(\overline{\text{RPS}}) = 0.10$. Hence, as regards the logarithmic score there is some evidence that the PM model performs significantly better.

We finally discuss possible reasons for the differences in the mean scores of models PM_7 and H_7^3. For a closer investigation, we looked at the mean scores stratified by the number of actually observed counts per region. For each score we observed the same

Table 4: The estimated parameters $\hat{\lambda}$ and $\hat{\rho}$ for models PM_2, PM_7, H_2^3 and H_7^3 for the Coxiellosis data.

Model	w_{ji}	$\hat{\lambda}$ (95%-CI)	$\hat{\rho}$ (95%-CI)
PM_2	$i \sim j$	0.73 (0.58;0.87)	0.0845 (0.0484;0.1229)
PM_7	$(CT_{ji}/n_j) \cdot b_j$	0.78 (0.67;0.83)	0.0048 (0.0028;0.0072)
H_2^3	$i \sim j$	0.38 (0.24;0.60)	0.0498 (0.0283;0.0876)
H_7^3	$(CT_{ji}/n_j) \cdot b_j$	0.36 (0.23;0.56)	0.0302 (0.0193;0.0470)

pattern: There are no cases for 150 regions and the difference between the PM and H^3 scores is essentially 0. Hence, only the remaining 35 regions with 1 to 6 observed cases cause the differences in the mean scores. The larger the observed count, the larger is the difference between the PM and H^3 scores. Hence, the PM model seems to perform better in predicting higher counts.
A different behavior of the PM and H^3 models can be observed, when iid ($\boldsymbol{\nu}$) or ICAR random effects ($\boldsymbol{\psi}$) are included on Stage 1 of model (6) or into η_{it} of model (7), respectively. For the PM models in our case study such iid or ICAR random effects are estimated to be zero and the obtained scores from the Table 3 do change only minimally. However, if one such random effect is included into an H^3 model and the weights w_{ji} are chosen as in, e.g., model H_7^3, the parameters $\hat{\lambda}$ and $\hat{\rho}$ are estimated to be zero and all variability between the regions is explained by the spatial random effect $\boldsymbol{\nu}$ or $\boldsymbol{\psi}$. The predictive scores obtained for models with an additional random effect are higher than for models including a weight scheme but no random effects (e.g. for inclusion of $\boldsymbol{\nu}$: $\overline{\text{SES}} = 0.607$, $\overline{\text{logS}} = 0.603$ and $\overline{\text{RPS}} = 0.239$, cf. Table 3). This behavior might be due to the fact that the data are very sparse. Hence, there are only a few yearly transitions with observed cases for both successive years present in the data, which are needed to estimate the parameters λ and ρ in the H^3 framework. So, if random effects are included additionally, they dominate the estimation process.
For completeness the estimated autoregressive parameters for some of the competing models are displayed in Table 4. In the PM context the estimates $\hat{\lambda}$ are generally larger than for the H^3 models. Note that the $\hat{\rho}$'s cannot be compared directly, since the weights are of a different absolute size.

5 Discussion

It was shown in this paper, how parameter- and observations-driven models for time series of infectious disease counts can be applied to network data using ready-to-use software. An emphasis is on the usage of the INLA method for approximate Bayesian inference in parameter-driven models, which can be conducted using a toolbox for generalized dynamic models.
The proposed parameter-driven models can be compared to so-called H^3 models on the basis of proper scoring rules, which are a tool to evaluate the predictive performance in terms of one-step-ahead predictions. In both case studies the parameter-driven models give better mean scores. Furthermore, the PM models tend to give better scores, if the observed number of cases is relatively large. A possible reason for this is their greater flexibility due to the complex hierarchical model structure. Unfortunately, the difference in model complexity between both approaches can hardly be

quantified. A disadvantage of the parameter-driven models discussed here is that the autoregressive parameters cannot be modelled as random effects and, hence, are the same for all regions in the study area. This restriction is not made by the H^3 models. For the Coxiellosis data we found that models assuming a spatial spread between neighbouring regions or by cattle trade exhibit a better predictive performance than models using an autoregressive component within a region only. The spatial spread is modelled best including the relative number of traded cattle weighted by the mean herd size in the origin region of the traded cattle as spread network. Hence, the cattle trade slightly outperforms a purely local, adjacency-driven spread mechanism. The mean herd size in a region elevates the risk of a spatial spread of the disease by cattle trade.
For future work it is desirable to estimate a data-driven transformation of the cattle trade weights. Furthermore, the results might be more precise, if time-varying cattle trade data are available. As an additional issue, Nöremark et al. (2011) state that between-country comparisons or the inclusion of data from adjacent countries can be helpful to parametrize disease spread models. Unfortunately, there are often differences in the data structure between national registries.

Acknowledgements

Financial support by the Swiss Federal Veterinary Office (BVET) is gratefully acknowledged. Special thanks to my colleague Michaela Paul for sharing her expertise about H^3 models.

References

Bigras-Poulin, M., Barfod, K., Mortensen, S., and Greiner, M. (2007). Relationsship of trade patterns of the Danish swine industry animal movements network to potential disease spread. *Preventive Veterinary Medicine* **80,** 143–165.

Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society - Series B* **63,** 823–841.

Brownstein, J. S., Wolfe, C. J., and Mandl, K. D. (2006). Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine* **3,** e401.

Chib, S., Greenberg, E., and Winkelmann, R. (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics* **86,** 33–54.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments (with discussion and reply). *Scandinavian Journal of Statistics* **8,** 93–115.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics* **65,** 1254–1261.

Diggle, P. J., Knorr-Held, L., Rowlingson, B., Su, T., Hawtin, P., and Bryant, T. N. (2004). On-line monitoring of public health surveillance data. In Brookmeyer, R. and Stroup, D., editors, *Monitoring the health of populations*. Oxford University Press.

Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001). The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science* **292,** 1155–1160.

Fèvre, E. M., de C. Bronsvoort, B. M., Hamilton, K. A., and Cleaveland, S. (2006). Animal movements and the spread of infectious diseases. *Trends in Microbiology* **14,** 125–131.

Frühwirth-Schnatter, S., Frühwirth, R., Held, L., and Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing* **19,** 479–492.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state-space modelling. *Biometrika* **93,** 827–841.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika* **85,** 215–227.

Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R., and Wint, W. (2005). Cattle movement and bovine tuberculosis in Great Britain. *Nature* **435,** 491–496.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102,** 359–378.

Green, D. M., Kiss, I. Z., and Kao, R. R. (2006). Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences* **273,** 2729–2735.

Guttorp, P. (1995). *Stochastic modeling of scientific data*. Chapman & Hall, London.

Harvey, A. C. (1981). *Time series models*. Phillip Alan, Oxford.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. University Press, Cambridge.

Hay, J. L. and Pettitt, A. N. (2001). Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* **2,** 433–444.

Held, L., Höhle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling* **5,** 187–199.

Held, L., Rufibach, K., and Balabdaoui, F. (2010). A score regression approach to assess calibration of continuous probabilistic predictions. *Biometrics* **66,** 1295–1305.

Hufnagel, L., Brockman, D., and Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 15124–15129.

Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis* **4,** 465–496.

Knorr-Held, L. and Richardson, S. (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics* **5,** 169–183.

Li, W. (1994). Time series models based on generalized linear models: some further results. *Biometrics* **50,** 506–511.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, Berlin.

Merler, S. and Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological sciences* **277,** 557–565.

Mugglin, A. S., Cressie, N., and Gemmell, I. (2002). Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine* **21,** 2703–2721.

Natale, F., Armando, G., Savini, L., Palma, D., Possenti, L., Fiore, G., and Calstri, P. (2009). Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine* **92,** 341–350.

Nelson, K. P. and Leroux, B. G. (2006). Statistical models for autocorrelated data. *Statistics in Medicine* **25,** 1413–1430.

Nöremark, M., Håkansson, N., Sternberg Lewerin, S., Lindberg, A., and Jonsson, A. (2011). Network analysis of cattle and pig movements in Sweden: Measures relevant for disease control and risk based surveillance. *Preventive Veterinary Medicine* DOI: 10.106/j.preventmed.2010.12.009.

Paul, M. and Held, L. (2011). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine* **30,** 1118–1136.

Paul, M., Held, L., and Toschke, A. M. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine* **27,** 6250–6267.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine* **29,** 1325–1339.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Riebler, A., Held, L., and Rue, H. (2010). Correlated multivariate age-period-cohort models. Technical report, University of Zurich, Switzerland.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields*. Chapman & Hall/CRC, London.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society - Series B* **71,** 319–392.

Ruiz-Cardenas, R., Krainski, E. T., and Rue, H. (2010). Fitting dynamic models using integrated nested Laplace approximations - INLA. Technical report, NTNU Trondheim, Norway.

Schrödle, B., Held, L., Riebler, A., and Danuser, J. (2011). Using INLA for the evaluation of veterinary surveillance data from Switzerland: A case study. *Journal of the Royal Statistical Society - Series C* **60,** 261–279.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society - Series B* **64,** 583–639.

Sun, D. and Ni, S. (2004). Bayesian analysis of vector-autoregressive models with noninformative priors. *Journal of Statistical Planing and Inference* **121,** 291–309.

White, P., Frankena, K., O'Keeffe, J., More, S. J., and Martin, S. W. (2010). Predictors of the first between-herd animal movement for cattle born in 2002 in Ireland. *Preventive Veterinary Medicine* **97,** 264–269.

Zeger, S. L. (1988). A regression model for time series counts. *Biometrika* **75,** 621–629.

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics* **44,** 1019–1031.

Web-based Supplementary Materials for "Assessing the impact of network data on the spatio-temporal spread of infectious diseases"

Birgit Schrödle, Leonhard Held and Håvard Rue

1 Web Appendix A

Univariate time series

First it is shown, how model PM_3 from Section 3.2 is coded in `INLA`. For a description of the seasonal term see equation (2). The Salmonella agona data are available in the `R` package `surveillance`.

```
> ### see Section 3.2
> ### PM_3: autoregression - yes, seasonality - yes
> # load salmonella agona data
> library(INLA)
> library(surveillance)
> data(salmonella.agona)
> Y <- salmonella.agona$observed
> T <- length(Y)
> t <- 1:T
> p <- 52
> sinterm <- sin(2*pi*t/p)
> costerm <- cos(2*pi*t/p)
> zeta <- t
> #
> # define model formula
> f_pm3 <- Y~1+f(zeta,model="ar1",param=c(0.1,0.001,0,0.2))+t+sinterm+costerm
> #
> # call model
> m_pm3 <- inla(f_pm3,family="Poisson",data=data.frame(Y,zeta,t,sinterm,costerm))
```

Multivariate time series

The implementation of the multivariate time series model (4) in `R` is similar to the univariate case. See as an application the code for model PM_1 for the Coxiellosis data (see Section 4.2). Here, the important feature is the `replicate`-option, which fits identical copies of a temporal AR(1) process for all regions.

```
> # Coxiellosis data are stored in a data.frame data_cox<-cbind(Y,E)
> # the ordering of the data is (i,t)
> Y <- data_cox[,1]
> E <- data_cox[,2]  # offset m_it
> #
> I <- 185
> T <- 6
> #
> zeta <- rep(1:T,I)
> id <- rep(1:I,each=T)
> #
> # define model formula for model: PM_3
> f_pm3 <- Y~f(zeta,model="ar1",replicate=id,param=c(0.1,0.001,0,0.2))
> #
> # call model
> m_pm3 <- inla(f_pm3,family="Poisson",E=E,data=data.frame(Y,E,zeta,id))
```

Multivariate time series models including a ρ component

If a ρ component as in equation (6) should be included, an augmented model containing pseudo-observations must be implemented. A detailed description of the available toolbox is given in Ruiz-Cardenas et al. (2010). To code the variable and weight vectors for the first and second stage of model (6) an automatic `coding` function can be used.

```
> # function to code the variable and weight vectors/matrices
> # of the components of the first and second stage of model (6)
> # the matrix wm contains the weights (w_ji)^T
> #
> coding<-function(I,T,wm){
+ # function to transform the (i,t) coordinates of an observation in a vector coordinate j
+ idx <- function(t,i,T){
+ j <- (i-1)*T+t
+ if (t<1) return(NA)
+ else return(j)
+ }
+ #
+ # initialize the second part of the variable vector for zeta and zetatm1 (lambda)
+ zeta.p2 <- rep(NA,I*T)
+ zetatm1.p2 <- rep(NA,I*T)
+ #
+ # initialize the second part of the rho component matrix and the respective weights
+ # the rho matrix contains one column for each region
+ rho.p2 <- matrix(NA,nrow=I*T,ncol=I)
+ w.rho.p2 <- matrix(NA,nrow=I*T,ncol=I)
+ #
+ # go through the vector with dimension I*T=J
+ j <- 1
+ for (i in 1:I){
+ for (t in 1:T){
+ # define the coordinates of zeta and zetatm1 for each vector node j
+ zeta.p2[j] <- idx(t,i,T)
+ zetatm1.p2[j] <- idx(t-1,i,T)
+ #
+ # define the rho coordinate and weight of each region k for vector node j
+ for (k in 1:I){
+ rho.p2[j,k] <- ifelse(wm[i,k]==0,NA,idx(t-1,k,T))
+ w.rho.p2[j,k] <- ifelse(wm[i,k]==0,NA,-wm[i,k])
+ }
+ j <- j+1}}
+ #
+ # put together the first and second part of the variable/weight vectors and matrices
+ zeta <- c(zeta.p2,zeta.p2)
+ zetatm1 <- c(rep(NA,I*T),zetatm1.p2)
+ rho <- rbind(matrix(NA,nrow=(I*T),ncol=I),rho.p2)
+ w.rho <- rbind(matrix(NA,nrow=(I*T),ncol=I),w.rho.p2)
+ #
+ return(list(zeta,zetatm1,rho,w.rho))}
```

The augmented model can be fitted using two different likelihoods for the real (Poisson) and pseudo-observations (Gaussian). The important feature here is the `same.as`-option, which allows to have the same ρ for each region.

```
> # set up the Y matrix (2*I*Tx2) with pseudo-observations for the evolution equation
> Y <- matrix(NA,nrow=(T+T)*I,2)
> Y[1:(I*T),1] <- data_cox[,1]
> Y[((I*T)+1):(2*T*I),2] <- 0
> #
> # code the variable vector for components on the first stage - here: alpha
```

```
> alpha <- c(rep(1,T*I),rep(NA,T*I))
> #
> # code the variable vectors for components on the first/second stage - here: zeta, zetatm1, rho
> # run function coding - wm can be a neighbourhood or other weight matrix
> code2ndstage <- coding(I,T,wm)
> # variable vector for zeta
> zeta <- code2ndstage[[1]]
> # variable vector for zetatm1
> zetatm1 <- code2ndstage[[2]]
> # weights for zetatm1
> w.zetatm1 <- ifelse(is.na(zetatm1)==TRUE,NA,-1)
> # variable matrix for rho
> rho <- code2ndstage[[3]]
> # weight matrix for rho
> w.rho <- code2ndstage[[4]]
> #
> # initialize the model formula
> # use the "copy" feature to obtain an identical copy of the zeta process
> # to estimate scaling parameters lambda (for zetatm1) and rho (for rho.1, rho.2,...)
> # set the option fixed=FALSE
> f <- "Y~f(zeta,model=\"iid\",fixed=TRUE,initial=-10)+
+                       +f(zetatm1,w.zetatm1,copy=\"zeta\",fixed=FALSE,param=c(0,1))+
+                       +f(rho.1,w.rho.1,copy=\"zeta\",fixed=FALSE,param=c(0,1))+alpha-1"
> #
> # or for a Fisher z-transformed lambda and rho (range-option)
> # f <- "Y~f(zeta,model=\"iid\",fixed=TRUE,initial=-10)+
> #         +f(zetatm1,w.zetatm1,copy=\"zeta\",fixed=FALSE,param=c(0,0.2),range=c(-1,1))+
> #         +f(rho.1,w.rho.1,copy=\"zeta\",fixed=FALSE,param=c(0,0.2),range=c(-1,1))+alpha-1"
> #
> # split the weight matrix for rho in separate vectors for each region - here: region 1
> vn <- paste("rho",1,sep=".")
> assign(vn,rho[,1])
> vn.w<-paste("w.rho",1,sep=".")
> assign(vn.w,w.rho[,1])
> #
> # run a split loop over all columns (regions) of the variable and weight matrix of rho
> for (i in 2:I){
+ vn <- paste("rho",i,sep=".")
+ assign(vn,rho[,i])
+ vn.w <- paste("w.rho",i,sep=".")
+ assign(vn.w,w.rho[,i])
+ #
+ # add a rho term for each region to the formula object
+ # to make sure that it's the same rho for each region, use option "same.as"
+ f <- paste(f,"+f(rho.",i,",w.rho.",i,",copy=\"zeta\",same.as=\"rho.1\")",sep="")}
> #
> # convert the formula in a formula object
> f_pm4 <- as.formula(f)
> #
> # define the offset E - on the first stage
> E <- c(data_cox[,2],rep(NA,I*T))
> #
> # call model - with two different likelihoods for the data and the pseudo-observations
> m_pm4 <- inla(f_pm4,family=c("Poisson","Gaussian"),data=data.frame(Y),
+               E=E,control.data=list(list(),list(initial=0)))
```

References

Ruiz-Cardenas, R., Krainski, E. T., and Rue, H. (2010). Fitting dynamic models using integrated nested Laplace approximations - INLA. Technical report, NTNU, Norway.

APPENDIX A

Analyzing veterinary surveillance data: Approaches to model the relationship between disease incidence and cattle trade

Birgit Schrödle, Leonhard Held & Michaela Paul

Extended abstract published in the Proceedings of the 25^{th} *International Workshop on Statistical Modelling, Glasgow, UK, 2010.*

Analyzing veterinary surveillance data: Approaches to model the relationship between disease incidence and cattle trade

Birgit Schrödle[1], Leonhard Held[1] and Michaela Paul[1]

[1] University of Zurich, Institute for Social and Preventive Medicine, Hirschengraben 84, 8001 Zurich, Switzerland, Email:birgit.schroedle@ifspm.uzh.ch

Abstract: Two approaches to the analysis of registry data for bovine diseases with regard to the relationship between disease incidence and cattle trade are proposed. Firstly, a parameter-driven spatio-temporal disease mapping model formulated in a hierarchical Bayesian framework is used. Various cattle movement parameters, e.g. the number and proportion of in-movements from infected regions, can be included as potential covariates. Within this context problems of such an endogenous covariate are discussed. Since a purely parameter-driven approach is often not adequate to depict local epidemics, a so-called observation-driven infectious disease model is proposed as a second possibility. It includes an autoregressive part for counts in the region of interest in the past. Additionally, the sum of previous cases in other regions weighted by cattle movements is added to assess the spread of the disease by trading. Both models are applied to cases of Coxiellosis in Switzerland, 2005 to 2009.

Keywords: Cattle trade; Spatio-temporal disease mapping; Infectious disease; INLA; Observation-driven

1 Introduction

The spread of a bovine disease can take place over short distances between adjacent or nearby farms borne by wind or insects, which typically results in local clustering of cases. However, disease dispersal also takes place over long distances caused by trade of infectious animals (Gilbert et al., 2005). Hence, the inclusion of cattle trade in an analysis of veterinary surveillance data might give hints towards the association of animal movement and disease presence.

As a first approach disease counts can be analyzed using a disease mapping model that considers spatial, temporal, and spatio-temporal trends (Knorr-Held, 2000; Schrödle and Held, 2009). Additionally, various cattle movement parameters can be included in the model using ecological regression (Clayton et al., 1993). Bayesian inference is conducted using integrated nested Laplace approximations (INLA), which was recently proposed in Rue et al. (2009).

As an alternative to this purely parameter-driven approach the spread of a bovine disease by cattle trade can be modelled within a likelihood-based infectious disease framework (Paul et al., 2008). Here, an autoregressive term for past counts in the region of interest and the sum of past cases in other regions weighted by cattle movements are part of the model formulation. The advantage of this so-called observation-driven model is that it is able to describe local epidemics.
Coxiellosis in cattle is an infectious, bacterial disease among ruminant animals, which can be spread by airborne infection. It can be the reason for an abortion, even in a late phase of the pregnancy. Cases of Coxiellosis were reported to the Swiss Federal Veterinary Office (BVET) between 2005 and 2009 and are available aggregated for 184 Swiss regions and the Principality of Liechtenstein on a yearly basis. The number of herds m_i in each region i is known.
Since 2008 it has been mandatory for Swiss stock-keepers to notify the BVET of all cattle movements. As the spatial pattern of movements is similar for 2008 and 2009 the considered movement parameters are assumed to be consistent from year to year.

2 Spatio-temporal disease mapping

For modelling spatio-temporal disease counts $y_{it} \sim \text{Po}(\exp(\eta_{it}))$ a nonparametric hierarchical Bayesian setting as proposed in Knorr-Held (2000) is used. The respective linear predictor can be written as

$$\eta_{it} = \log(m_i) + \xi + \nu_i + \psi_i + \gamma_t + \phi_t + \delta_{it}, \tag{1}$$

where ξ is an intercept, ν_i and ψ_i are spatially unstructured and structured effects and γ_t and ϕ_t are temporal main effects, specified as an i.i.d. term and a random walk of first order, respectively. The term δ_{it} accounts for spatio-temporal interaction and can be specified assuming four different types of interaction between time and space (Knorr-Held, 2000). To account for covariates x_{it} as introduced in the following paragraph (1) can be extended to

$$\eta_{it} = \log(m_i) + \xi + \nu_i + \psi_i + \gamma_t + \phi_t + \delta_{it} + \beta \cdot x_{it}. \tag{2}$$

The total number of in-movements (model acronym: TOT) and the absolute number (A) and proportion (P) of in-movements from regions with elevated risk are considered as potentially associated with disease presence. Since we assume that the movement pattern does not change from year to year the total number of in-movements is constant over time. The movements from infected regions are time-varying and defined using a two-stage process: A separate spatial disease mapping model is fitted for each year, including only ν_i and ψ_i from (1) (Besag et al., 1991). Regions are indicated

as infected when exceeding two different thresholds, namely an estimated relative risk larger than 2 and 3, respectively. At the second stage the models are fitted using a time lag of one and two years, respectively, to detect the incubation period of the disease (TL1 and TL2). As the time-varying covariate is derived using previous observations it is a so-called endogenous or feedback variable. This issue will be discussed briefly in Section 4.
All models are fit using integrated nested Laplace approximations (INLA). This approach for approximate Bayesian inference was recently proposed by Rue et al. (2009) as an alternative to Markov chain Monte Carlo mechanisms. The advantage of INLA is that it runs in remarkably fast computational time and returns accurate parameter estimates for a wide range of models. Additionally, INLA computes the deviance information criterion (DIC) as tool for Bayesian model choice. All analyses in this paper were conducted using the `R INLA` package build on the 1st of February 2010, version 1.668.

3 Infectious disease model

A purely parameter-driven model as proposed in Section 2 might not be able to describe localized epidemics which can often be found in veterinary disease surveillance data (Held et al., 2005). Hence, a so-called observation-driven model is built including the number of cases $y_{i,t-1}$ in the past. In its simplest formulation the observations y_{it} are Poisson distributed with mean

$$\mu_{it} = \lambda \cdot y_{i,t-1} + m_i \cdot \exp(\alpha) \tag{3}$$

and $\lambda > 0$ (Held et al., 2005). The parameter α accounts for all residual variation. Cases at times $t - k$, $k > 1$, could be considered as well.
As an addition, the sum of counts in all other regions j weighted by a factor w_{ji} can be added to model the spatial spread of the disease over time. In Paul et al. (2008) the respective mean is specified as

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha) \tag{4}$$

with $\lambda, \rho > 0$. To assess the association between cattle movement and disease presence the square root of the absolute number of cattle movements (CM) between regions j and i are used as weights w_{ji} in this application. Other weights w_{ji} can also be considered (Paul et al., 2008). Here, models with $w_{ji} = 1$ for all j and for all $j \sim i$, respectively, are fit as alternatives. The term $j \sim i$ denotes all regions j which are neighbours of region i.
The parameter α in (4) can be split into an intercept and a linear time trend

$$\mu_{it} = \lambda \cdot y_{i,t-1} + \rho \cdot \sum_{j \neq i} w_{ji} \cdot y_{j,t-1} + m_i \cdot \exp(\alpha + \zeta \cdot t). \tag{5}$$

TABLE 1. Spatio-temporal disease mapping (see Section 2): The DIC and the posterior mean of the respective cattle trade parameter along with its 95%-credible interval are shown for each model. For the model without covariate (1) the DIC is shown only.

		RR> 2		RR> 3	
		DIC	$\hat{\beta}$	DIC	$\hat{\beta}$
TL1	A	939.3	$0.017\,[0.008; 0.027]$	942.7	$0.012\,[0.004; 0.020]$
	P	947.0	$1.29\,[0.39; 2.18]$	950.0	$0.96\,[-0.02; 1.94]$
TL2	A	945.2	$0.009\,[-0.000; 0.018]$	951.0	$0.007\,[-0.003; 0.016]$
	P	948.0	$0.83\,[-0.038; 1.69]$	951.1	$0.85\,[-0.20; 1.90]$
TOT		955.8	$0.005\,[-0.006; 0.017]$		
(1)		954.6			

Extensions for a region-specific random effect are also possible (Paul et al., 2009), but the computation of the results might suffer from numerical problems if the number of regions is large. Hence, they are not considered here.
Maximum likelihood inference is performed using iterative algorithms as described in Held et al. (2005) and Paul et al. (2008). The AIC is calculated for model choice.

4 Results

Regarding spatio-temporal disease mapping, model (1) was run without covariate for all four possible types of space-time interaction. The model including an interaction term of Type II had the lowest DIC and was chosen as basis model for the ecological regression including cattle trade quantities. All results are summarized in Table 1. The DIC is lowest for the model including the absolute number of in-movements from infected regions with relative risk larger than 2 and a time lag of one year. This is plausible considering the nature of the disease. Models including the absolute number of in-movements are generally preferred compared to models involving the proportion of cattle trade from infected regions. A positive association is obtained for all covariates. For three of the models with a time lag of one year the 95%-credible interval includes only positive values.
As noted in Section 2 the number of in-movements from infected regions is an endogenous covariate. If the pattern of the disease exhibits local clusters, the infected areas chosen by the two-stage process typically are a few groups of neighboring regions. As cattle trade is much larger between neighboring regions, the respective parameters might just explain parts of the local spatial clustering of cases in the data. One hint pointing in this direction is the fact that the estimated variance of the spatially structured effect ψ_i in

TABLE 2. Infectious disease model (see Section 3): The AIC and the estimated coefficients along with their standard errors are shown for each model.

	w_{ji}	AIC	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\zeta}$
(3)	–	1092.8	0.44 (0.05)		
(4)	1	1094.8	0.44 (0.05)	0.0000 (0.0000)	
(4)	1, if $j \sim i$	1062.7	0.43 (0.05)	0.0491 (0.0103)	
(4)	$\sqrt{\text{CM}}$	1082.0	0.44 (0.05)	0.0005 (0.0002)	
(5)	$\sqrt{\text{CM}}$	1083.5	0.44 (0.05)	0.0005 (0.0002)	0.08 (0.11)

(1) drops after inclusion of cattle movement in the model. Unfortunately, it cannot be quantified to what extent such confounding is present.
Results for the infectious disease models are shown in Table 2. With regard to AIC model (4) using $\sqrt{\text{CM}}$ as weights performs better than model (3). Hence, the autoregressive inclusion of counts from other regions weighted by cattle trade provides a better fit. The respective parameter estimate $\hat{\rho}$ is positive (0.0005) with a small standard error (0.0002). In contrast, the alternative model with $w_{ji} = 1$ for all j is not better than (3). If the counts in neighbouring regions $j \sim i$ are considered as additional explanatory variables the AIC is even smaller than for the model including cattle trade and the estimated coefficient is significantly positive (0.0491). Hence, a high local clustering of cases is present. For (5) a positive linear time trend is estimated (0.08), but it is not significantly different from zero.

5 Discussion

Two very different approaches were applied to data on Coxiellosis in cattle to assess the spread of the disease by cattle trade. In both cases model choice criteria and estimated coefficients indicate a positive association between animal movement and disease presence. Nevertheless, both approaches are not without problems. The disease mapping approach makes use of an endogenous covariate which might result in a confounding problem. With regard to the infectious disease model first steps in the direction of a mixture of the parameter- and observation-driven approach are taken when fitting (5). Nevertheless, it would be desirable to include spatial effects as well, especially a spatially structured effect to account for local clustering of the disease. Hence, it must be explored if both approaches could be combined by substituting α in (3) by (1) (without $(\log m_i)$). In this new setting parameter estimation might be possible in a Bayesian framework as algorithms used for maximum likelihood inference (Paul et al., 2008) will possibly suffer from numerical problems. Furthermore, the two approaches are not comparable at the moment as different model choice criteria are derived and different components are included.

Acknowledgments: Financial support by the Swiss Federal Veterinary Office (BVET) is gratefully acknowledged.

References

Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.

Clayton, D., Bernardinelli, L. and Montomoli, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology*, **22**, 1193-1202.

Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R. and Wint, W. (2005). Cattle movement and bovine tuberculosis in Great Britain. *Nature*, **435**, 491-496.

Held, L., Höhle, M. and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, **5**, 187-199.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19**, 2555-2568.

Paul, M., Held, L. and Toschke, A. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, **27**, 6250-6267.

Paul, M. and Held, L. (2009). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Technical Report, University of Zurich.*

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of the Royal Statistical Society, Series B*, **71**, 319-392.

Schrödle, B. and Held, L. (2009). Spatio-temporal disease mapping using INLA. *Technical Report, University of Zurich.*